8-16-2011

Optimizations and applications in head-mounted video-based eye tracking

Feng Li

Optimizations and Applications in Head-Mounted Video-Based Eye Tracking

by

Feng Li

M.S. Zhejiang University, 2003

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the Chester F. Carlson Center for Imaging Science
Rochester Institute of Technology

August 16, 2011

Signature of the Author _____

Accepted by _____

Coordinator, Ph.D. Degree Program **Date**

CHESTER F. CARLSON CENTER FOR IMAGING SCIENCE

ROCHESTER INSTITUTE OF TECHNOLOGY

ROCHESTER, NEW YORK

CERTIFICATE OF APPROVAL

Ph.D. DEGREE DISSERTATION

The Ph.D. Degree Dissertation of Feng Li
has been examined and approved by the
dissertation committee as satisfactory for the
dissertation required for the
Ph.D. degree in Imaging Science

Dr. Jeff B. Pelz, dissertation Advisor

Dr. Daniel R. Lawrence

Dr. Carl Salvaggio

Dr. Jinwei Gu

Date

DISSERTATION RELEASE PERMISSON

ROCHESTER INSTITUTE OF TECHNOLOGY

CHESTER F. CARLSON CENTER FOR IMAGING SCIENCE

Title of Dissertation:
Optimizations and Applications in Head-Mounted Video-Based Eye Tracking

I, Feng Li, hereby grant permission to Wallace Memorial Library of R.I.T. to reproduce my thesis in whole or in part. Any reproduction will not be for commercial use or profit.

Signature _____

Date

Optimizations and Applications in Head-Mounted Video-Based Eye Tracking

by
Feng Li

Submitted to the
Chester F. Carlson Center for Imaging Science
in partial fulfillment of the requirements
for the Doctor of Philosophy Degree
at the Rochester Institute of Technology

Abstract

Video-based eye tracking techniques have become increasingly attractive in many research fields, such as visual perception and human-computer interface design. The technique primarily relies on the positional difference between the center of the eye's pupil and the first-surface reflection at the cornea, the corneal reflection (CR). This difference vector is mapped to determine an observer's point of regard (POR). In current head-mounted video-based eye trackers, the systems are limited in several aspects, such as inadequate measurement range and misdetection of eye features (pupil and CR). This research first proposes a new 'structured illumination' configuration, using multiple IREDs to illuminate the eye, to ensure that eye positions can still be tracked even during extreme eye movements (up to ±45° horizontally and ±25° vertically). Then eye features are detected by a two-stage processing approach. First, potential CRs and the pupil are isolated based on statistical information in an eye image. Second, genuine CRs are distinguished by a novel CR location prediction technique based on the well-correlated

relationship between the offset of the pupil and that of the CR. The optical relationship of the pupil and CR offsets derived in this thesis can be applied to two typical illumination configurations – collimated and near-source ones- in the video-based eye tracking system. The relationships from the optical derivation and that from an experimental measurement match well.

Two application studies, smooth pursuit dynamics in controlled static (laboratory) and unconstrained vibrating (car) environments were conducted. In the first study, the extended stimuli (color photographs subtending 2° and 17°, respectively) were found to enhance smooth pursuit movements induced by realistic images, and the eye velocity for tracking a small dot (subtending <0.1°) was saturated at about 64 deg/sec while the saturation velocity occurred at higher velocities for the extended images. The difference in gain due to target size was significant between dot and the two extended stimuli, while no statistical difference existed between the two extended stimuli. In the second study, twovisual stimuli same as in the first study were used. The visual performance was impaired dramatically due to the whole body motion in the car, even in the tracking of a slowly moving target (2 deg/sec); the eye was found not able to perform a pursuit task as smooth as in the static environment though the unconstrained head motion in the unstable condition was supposed to enhance the visual performance.

Acknowledgements

This thesis would not be possible without the invaluable help and support from a numerous of people.

First and foremost, I would like to thank my thesis advisor Dr. Jeff B Pelz for his ability to motivate, explain complicated matters and constant support during every stage of this endeavor. I am fortunate to have had him as a mentor. His dedication towards his students is rare and greatly appreciated.

I wish to express my gratitude to Dr. Daniel R. Lawrence, Dr. Carl Salvaggio, and Dr. Jinwei Gu for serving on my thesis committee and reviewing the thesis.

Many people, current and past members of the Visual Perception Laboratory and Multidisciplinary Vision Research Laboratory at RIT, helped in the development of this work. I would like to think Dr. Andrew Herbert, Dr. Mitchell R. Rosen, Dr. Sunsan Munn, Marianne Lipps and Mary Ellen Arndt for their stimulating discussions about both the thesis work and a vast array of other interesting topics. I would like to thank Christopher Louten, Monica Cook, Nicholas MacDowell and other undergraduates in the lab for assisting me in data collection and other matters.

I want to extend my gratitude to the whole Chester F. Carlson Center for Imaging Science for providing a pleasant and inspiring atmosphere.

I am deeply indebted to my family for all their love and support throughout these graduate years. My parents and my sister have sacrificed so much to give me an opportunity to fulfill my dreams. My lovely daughters Emily and Angela brought me measureless happiness over the past two years. Most important of all, none of this would have been possible without the tireless love and support of my wife Nan. For this and much more, I dedicate this thesis to them.

Contents

List of Figures

List of Tables

Chapter 1 Introduction

Beginning as early as the nineteenth century, researchers have been eager to explore hidden mysteries behind the human visual system. Since then, many groundbreaking techniques for recording eye movements have been invented and applied in practice: in 1898, a mechanical method involved attaching a small cap to a "cocainized eye" (Delabarre, 1898); today techniques such as electro-oculography, scleral search coils, dual-Purkinje trackers and video-based methods provide ways to serve researchers' diverse demands in contemporary studies.

These eye tracking techniques are filled with tradeoffs in comfort, accuracy, noise, cost, ease of calibration, suitability to a large population, and so forth. None of them completely satisfies the diverse necessities of researchers. Because of its minimal obtrusiveness to observers, relatively easy set-up and reliance on rapidly developing optical and electronic imaging devices, video-based eye tracking has become one of the most popular eye-tracking techniques. Video-based eye tracking commonly relies on the positional difference between the pupil center and the first-surface corneal reflection of a light source (CR) to map the line of sight. There are still a number of challenging problems remaining in video-based eye tracking, such as track loss due to large eye movements and specular reflection artifacts being mistaken for the desired CR. In this

thesis, a new structured illumination configuration, utilizing an array of IREDs to illuminate the eye, is proposed, which provides reliable difference vectors even during extreme eye movements. This design, along with novel image processing algorithms for detecting and isolating the eye features, deals with the two aforementioned problems and also address other problematic issues in video-based eye tracking. An ideal design seeks to achieve a measurement range up to ±50° horizontally and ±40° vertically with an accuracy of 1° or better.

As the invention and advance of new technologies aid to explore certain phenomenon or solve some problems, the improvement in eye tracking techniques enable to understand the characteristics of eye movements. Smooth pursuit is one of the most important eye movements, which is the eye motion for stabilizing the image of small moving objects on the fovea, the highest visual acuity zone of the retina. By applying the video-based eye tracking technique, the dynamics of smooth pursuit eye movements in two very different conditions were studied in the thesis.

In the first study, which was conducted in a static condition (laboratory), color photographs of small (<0.1°) and extended targets (2° and 17°, respectively) were used to induce pursuit eye movements, while the head was stabilized to a chinrest. While the performance of the eye in the pursuing tasks became worse with the increase of the target velocity, the extended targets were found to enhance the pursuit execution of the eye in comparison with the small target. And the improvements by two extended targets were found to be statistically insignificant. The eye velocity limit in tracking the small target in the experiment was 64 deg/sec while the limits in tracking extended targets were higher.

In comparison with the controlled static condition in the first study, the second one was conducted in a moving vehicle where the head was allowed to move freely. Another

important difference between two studies was that the target was shown at a distance in the first study while it was moving close to the eye in the second study. The pursuit eye movements in the unstable condition were never as smooth as that in the static condition because the whole body vibration generated by the vehicle motion impaired the visual performance to a great degree.

The thesis is divided into 11 chapters: *Chapters 1* and *2* give an introduction and statement of the work. *Chapters 3* and *4* overview the physiological optics of the eye and the characteristics of different kinds of eye movements. *Chapter 5* surveys eye-tracking techniques and their applications in the study of eye movements in visual tasks. *Chapter 6* addresses some problems in current commercial and laboratory video-based eye tracking systems and proposes potential solutions. *Chapters 7* and *8* present results of deriving optical relationships of the movements of the eye features (pupil and CR), improving measurement range and robustness of the new system by structured illumination. Two applications of video-based eye tracking technique in the study of the dynamics of smooth pursuit are detailed in *Chapters 9 and 10*, respectively. Finally *Chapter 11* draws conclusions of the thesis work and proposes the direction of future work.

Chapter 2 Statement of Work

The primary objectives of this thesis are to (1) optimize head-mounted video-based eye tracking based on the RIT Lightweight Tracker, and (2) apply vide-based eye tracking techniques in the studies of smooth pursuit dynamics in variant conditions. Specifically, the following eight aspects were pursued.

1. Gained related background knowledge, including the optics of the eye, the dynamics of eye movements, and the mechanisms of visual and cognitive processes in visual tasks.
2. Studied historical and state-of-the-art eye tracking techniques and became familiar with the operations of some commercial and laboratory eye trackers (RIT, ISCAN, ASL and Dual-Purkinje) as well as their calibration routine and data analysis software.
3. Surveyed eye tracking algorithms in the literature, explored challenges in video-based eye tracking, and proposed corresponding solutions.
4. Built a new structured illumination configuration, prototyped the hardware (e.g., headgear, illuminator and circuitry), and proved the concept. The new design provided a measurement range up to ±40° horizontally and ±25° vertically.

5. Designed and implemented image processing algorithms for detecting and isolating the eye features. These algorithms were able to deal with some primary noise in the eye image, and output pupil and CR positions with a high accuracy.
6. Applied video-based eye tracking techniques in measuring the dynamics of smooth pursuit eye movements in a controlled static condition.
7. Studied the eye's pursuit performance in an unstable environment, where the head motion was unconstrained.
8. Drew conclusions and brought forward directions for future work.

Chapter 3 Optics of the Eye

The cornea, pupil, and lens are the primary optical components of the human eye (Figure 1). Rays penetrating through the cornea, the aqueous humor, the lens, and the vitreous humor, strike the photosensitive receptors in the retina, where optical signals are converted to electrochemical signals and then transmitted to the brain. Information about the optics of the eye is essential for understanding the eye tracking techniques and effectively applying the techniques in eye movement studies. The basics of the optical characteristics of eye components introduced in this chapter are also used to derive a model (Chapter 7) to improve the detection and isolation of eye features for a video-based eye tracker (Chapter 8).

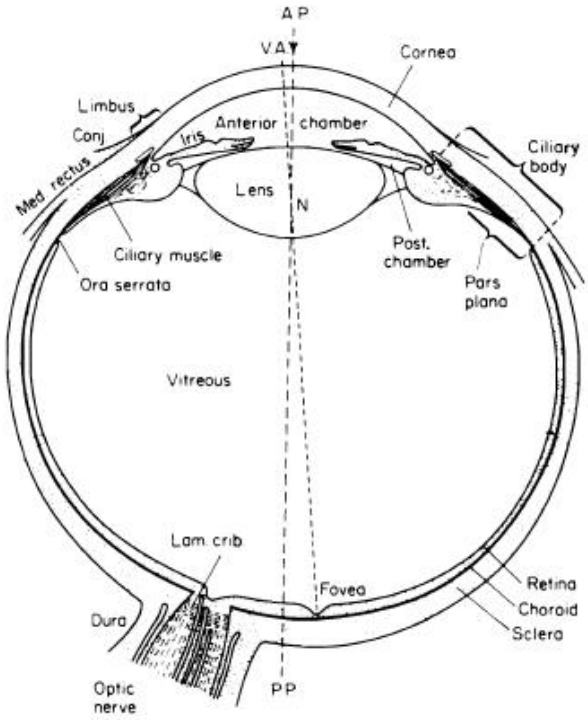

Figure 1. Sagittal horizontal section of an adult human eye. PP, posterior pole; AP, anterior pole; VA, visual axis. Courtesy of Davson, 1990, page 3.

3.1 The Cornea

The cornea is a tough, transparent membrane on the outer surface of the eye. Given the fact that there is a great change of refractive index, especially on the air-cornea interface, the cornea dominates (about 73%) the refracting power of the eye (Pedrotti & Pedrotti, 1998, page 195). The refractive index of the cornea depends on which part is measured, while its mean value is usually taken as 1.376 (Grand & Hage, 1980, pp 65-66).

The cornea is approximately spherical over the central 25 degrees of visual angle and flattens out beyond this region. A lot of studies have been done for modeling the

asphericity of the anterior surface of the cornea, among which 3D conicoid or 2D conic models are commonly used (Atchison & Smith, 2000).

The conicoid form is represented by the equation:

$$X^2 + Y^2 + (1+Q)Z^2 - 2ZR = 0 \tag{3-1}$$

where *X*, *Y,* and *Z* are values in the tangential, sagittal and optical axis direction, *R* is the vertex radius of the curvature, and *Q* is the asphericity having following definitions:

Q < -1: hyperboloid

Q = -1: paraboloid

-1 < Q < 0: ellipsoid, with the major axis in the Z-axis

Q = 0: sphere

Q > 0: ellipsoid, with the major axis in the X-Y plane.

The conic model has the following form:

$$\frac{(Z-a)^2}{a^2} + \frac{Y^2}{b^2} = 1 \tag{3-2}$$

where *a* and *b* are the semi-lengths of the ellipse. The eccentricity *e* is used to define the shape of the ellipse, which satisfies the following equation if taking the Z-axis as the major axis:

$$e^2 = 1 - \frac{b^2}{a^2} \tag{3-3}$$

Equations (3-1) and (3-2) are associated with each other by the relationships:

$$R = b^2 / a \tag{3-4}$$

$$Q = b^2 / a^2 - 1 \tag{3-5}$$

The posterior surface of the cornea is less interesting because of the difficulty of measurement and relatively small refracting power at the interface between the cornea and aqueous humor, which has a refractive index of about 1.336 ($n_{water} \approx 1.333$).

3.2 The Iris and Pupil

The iris is the aperture stop of the eye, functioning similarly to that in consumer cameras. Two muscles connecting to the iris control the dilation and constriction of the pupil, which is the circular opening in the middle of the iris that limits the amount of light entering the eye. Aside from controlling the amount of light that enters the eye, the constriction of the iris reduces the optical aberration by limiting the coming light; people often squint the eye to produce a sharp image on the fovea. Brown, blue, and green are the primary colors of the iris (or the eye). The eye color is determined by the concentration and distribution of pigmentation in the iris fibers; lightly pigmented irises appearing blue and heavily pigmented ones look brown (Daugman, 2004).

The actual pupil center lies on the frontal vertex of the lens, about 3.6mm away from the corneal vertex. When we look into someone's eye, what we see is not the actual pupil but an apparent or entrance pupil, which is the image of the actual pupil formed by the light refracted by the cornea. In the image formation process, the cornea acts as a concave surface diverging the rays (Figure 2 left).

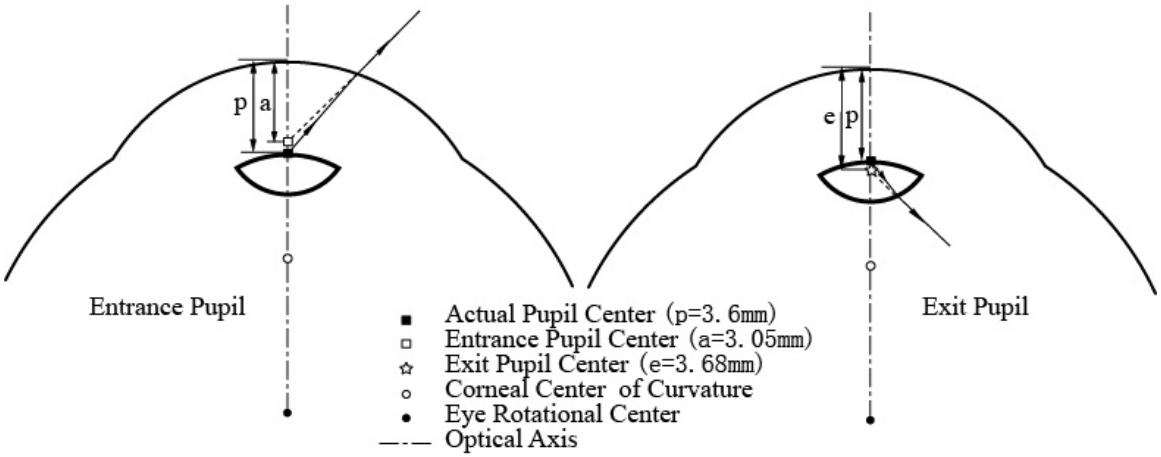

Figure 2. Formation of the entrance and exit pupil.

Based on the Gullstrand Number 2 Simplified Eye (see Section 3.6.2), the size and position of the entrance pupil can be determined by the following equation:

$$\frac{n'}{p}+\frac{n}{a}=\frac{n-n'}{r}\Rightarrow a=3.05\,\text{mm} \tag{3-6}$$

where n=1 and n′=1.336 are the refractive index of the air and aqueous humor, *r* is the radius of curvature of the cornea, *p=-3.6mm* (sign convention see Hecht, 1998) and *a* are the distances of the actual and entrance pupil with respect to the vertex of the cornea, respectively. Note that the apparent pupil center always falls on the optical axis of the eye, no matter whether the eye translates or rotates, if the actual pupil center and rotational center of the eye are assumed to be both on the optical axis of the eye. The position of the pupil is used to derive the optical relationship of the pupil and the first surface reflection of the cornea in Chapter 7.

Transverse magnification is given by:

$$M=\frac{n'/p}{n/a}=\frac{1.336/3.60}{1/3.05}=1.13 \tag{3-7}$$

In the case of an actual pupil size of 3mm, the entrance pupil would be 3×1.13=3.39mm. So the apparent pupil is located closer to the cornea and is slightly larger than the actual pupil.

The exit pupil is the image formed by the optical elements behind the actual pupil (Figure 2 right). It is situated about 0.08mm behind the actual pupil and is 3.1 percent larger.

The relative positions and sizes of the actual, entrance, and exit pupil of the eye are shown in Figure 3.

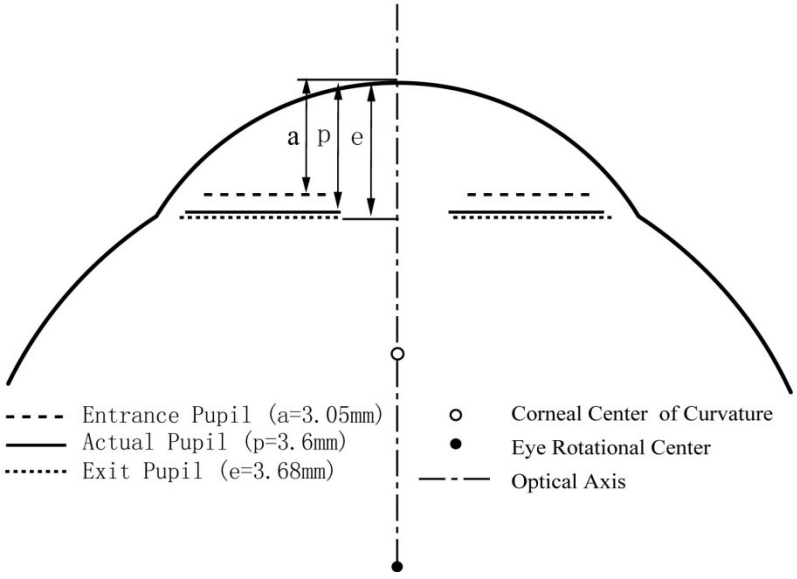

Figure 3. The relative positions and sizes of the actual, entrance and exit pupil.

Because the eye is not a rotationally symmetric optical system, the pupil is not centered in the optical axis of the eye, but displaced nasally by about 0.5 mm (Atchison & Smith, 2000). Furthermore, the pupil center will shift a small amount with the constriction and dilation of the pupil, up to 0.3mm-0.6mm according to different studies

in the literature (Walsh, 1988; Wilson, Campbell et al., 1992; Yang, Thompson et al., 2002).

The pupil diameter typically varies between 2mm to 8mm, resulting in an illumination ratio on the retina of 1:16 (proportional to the pupil area). In the early days, researchers suspected that the automatic adjustment of the pupil diameter was to keep an approximately constant energy on the retina (Reeves, 1920). However, recent studies (Atchison & Smith, 2000) argued this hypothesis because the change of the pupil area is only by a factor of 16 while the luminance change the eye may be exposed to is by over 10^5 times (from full moonlight $\approx 0.01 cd/m^2$ to bright daylight $\approx 1000 cd/m^2$). Those researches gave a new explanation that the pupil size varies in order to adaptively optimize the visual acuity in different illumination levels.

Among a number of factors (e.g., viewing distance, mental status and age) affecting the pupil size, illumination is the most important one (Atchison & Smith, 2000) (Figure 4). The pupil responds with dilation in size to the illumination decrease and vice versa.

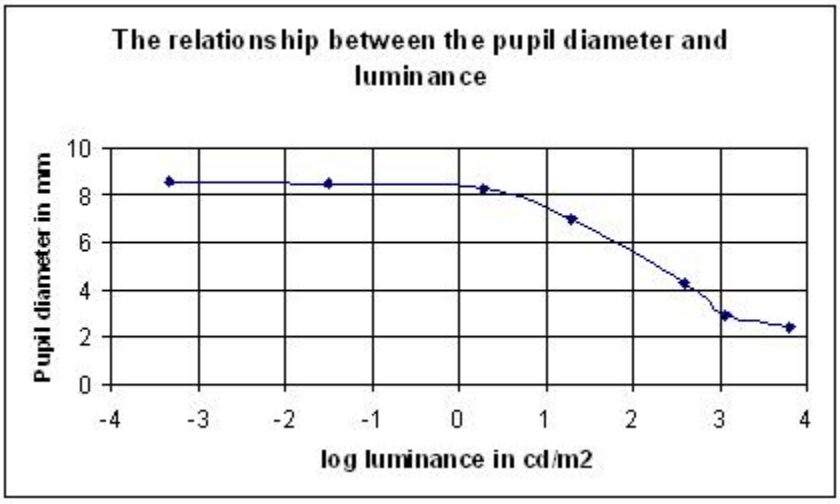

Figure 4. The relationship between the pupil diameter and illumination (Reeve, 1920). A typical office lighting is 120 cd/m2.

For an optical system, decreasing its aperture size reduces the aberration in the formed image (paraxial optics has a minimum aberration). Accordingly, a smaller pupil size improves the visual acuity by limiting the amount of light that enters the eye up to a point. It is known that a pupil diameter of 3mm is optimal for the visual acuity because diffraction effects become deteriorative when the pupil size becomes sufficiently small (Davson, 1990).

The location of the pupil and its size are important for many eye tracking apparatuses (e.g., video-based eye tracking introduced in Section 5.5). No matter what techniques are used to isolate and determine the pupil, for instance edge-based, area-based or others, a large pupil size is desirable since it provides a good "signal" and hence benefits the measurement. Especially for those video-based eye trackers relying on a bright-pupil effect (see Section 3.7.1), because the pupil brightness is directly proportional to the pupil area (Ebisawa, 1995; Karlene, Cindy et al., 2002), a larger pupil has a higher contrast with its surroundings (e.g., the iris), making further processing operations easier.

The pupil appears as a circular disk when it is observed along the direction of the optical axis of the eye. When the pupil is viewed obliquely its amplitude in the tangential direction becomes shorter while that in the sagittal direction remains unchanged. The eccentricity of the elliptical shape (Figure 5) of the pupil can be modeled by a simple geometrical relationship (Atchison & Smith, 2000):

$$Dt = D*cos\ \theta \tag{3-8}$$

$$Ds = D \tag{3-9}$$

where D is the actual diameter of the entrance pupil viewed along the optical axis, Dt is

the diameter in the tangential direction viewed along an oblique angle θ, and $Ds = D$ is the diameter in the sagittal direction keeps constant viewed along any oblique angle.

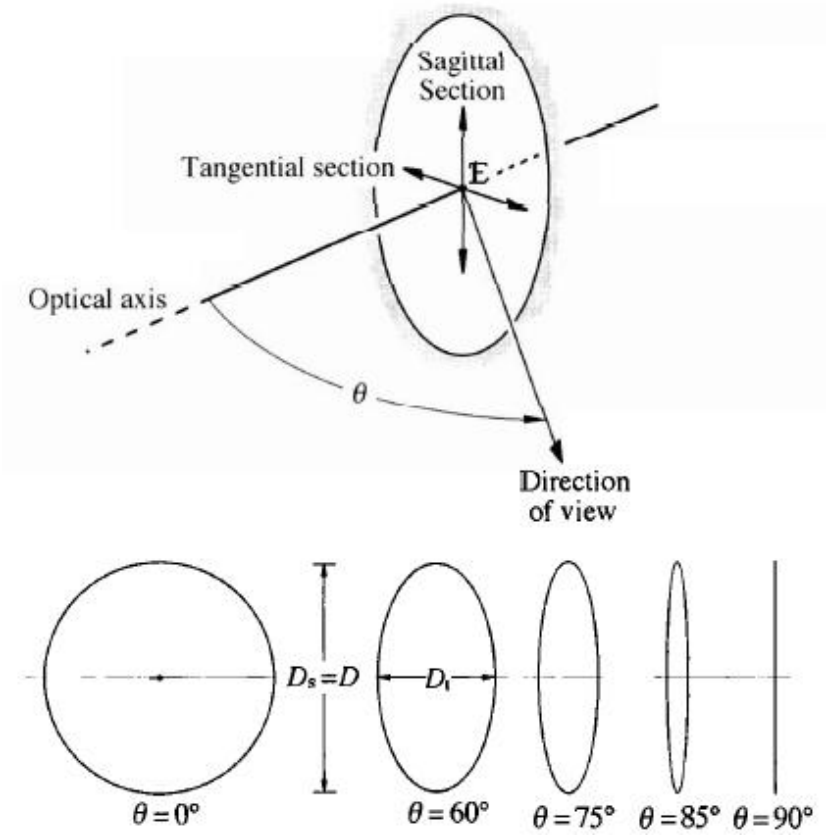

Figure 5. The pupil shape viewed obliquely. From Smith, 2003.

3.3 The Lens

Behind the pupil is another transparent structure called the lens (or crystalline lens). It has about 20% of the refractive power of the eye whereas it has the exclusive power of accommodation (focusing on objects at various distances). The adjustment is controlled by ciliary muscles attached (via zonules) to the periphery of the lens; the relaxation and

constriction of the muscles control the bulging and easing of the lens, respectively. During a constriction, first the muscles tighten, releasing the tension at the zonules, then the lens increases its curvature, allowing the eye to focus on a nearby object. The lens has a non-homogenous refractive index, changing from about 1.41 at the center to 1.39 at the rim (Pedrotti & Pedrotti, 1998). This gradient index structure serves to reduce spherical aberrations in the image formation process.

During and right after a saccade, the lens is found to have a displacement with respect to the optical axis of the eye (Deubel & Bridgeman, 1995a, 1995b). This dislocation causes the overestimation of saccadic eye movement (Section 4.2) by dual Purkinje eye trackers (Section 5.4.1), because one of the features used for tracking (P4) is the image formed by the back surface of the lens (see Section 5.4.1).

3.4 The Retina

The retina is the last stop of light in the eye, where optical information is converted into electro-chemical signals and then transmitted to the brain via optical nerves. This 0.5mm-thick membrane is composed of three layers (photoreceptor layer, inner neural layer and ganglion layer) of cell bodies separated by two plexiform layers (inner and outer layers) (Figure 6). The retina packs in a backward fashion to the pathway of incoming light: light penetrates through the ganglion layer, inner neural layer and then strikes the photoreceptor layer, which is the most outer layer of the retina; however, signal transmission from the retina to the brain works in a reverse direction, beginning from the photoreceptors.

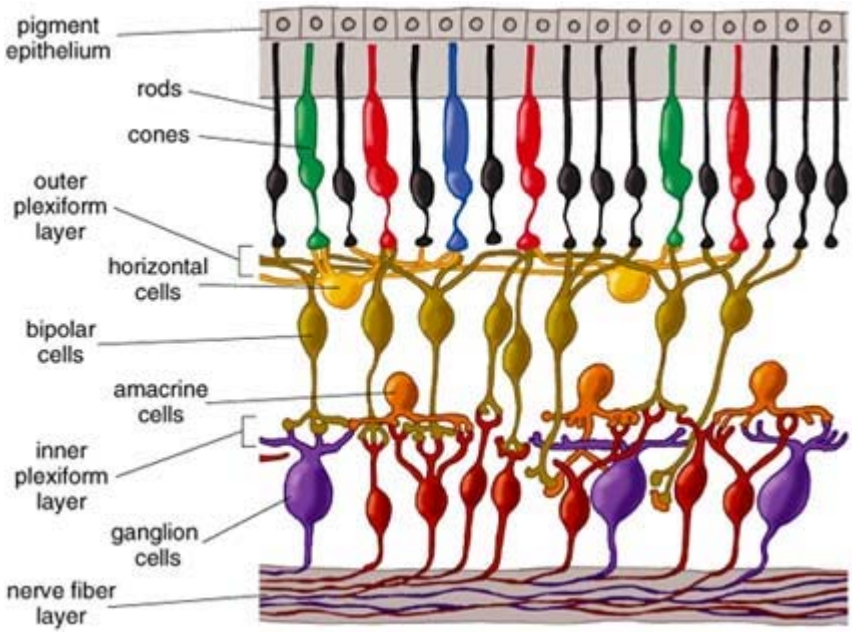

Figure 6. Structure of the retina. From Helga Kolb, 2003.

Two distinct types of photoreceptors (rods and cones) (Figure 7) endow the eye with great capabilities to work in different illumination conditions. The number of rods is about 15 times than that of cones, 100-120 million rods versus 7-8 million cones (Pedrotti & Pedrotti, 1998).

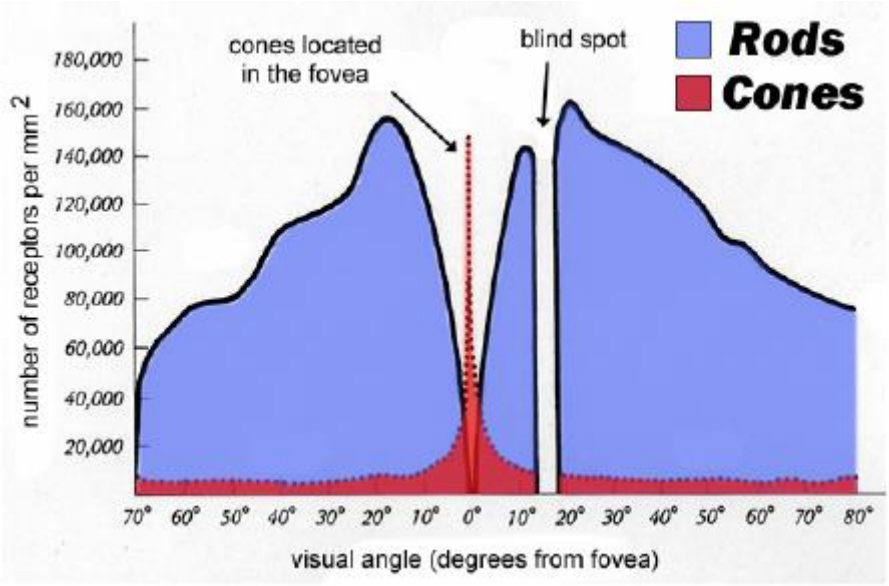

Figure 7. The distribution of the rods and cones in the retina (left eye). From Falk, Brill, and Stork, 1986, p. 153.

Cones cover the whole retina (except for the optical nerve), but are heavily concentrated in the fovea (a 1~2 degree area providing the best spatial and color vision). A single cone is connected to several ganglions cells such that it enables us to discern high spatial details, with the tradeoff of a relative low sensitivity. In common lighting conditions, the cones are the primary active sensors (photopic vision). Three classes of cones, L-cones (long wavelength), M-cones (medium wavelength), and S-cones (short wavelength), respond maximally to different wavelengths of light. This makes the cones exclusively capable of discerning color. Visual acuity and color vision decreases away from the fovea due to the drop of cone density.

Rods, another kind of photoreceptor, are distributed unevenly in the retina, peaking around 15~20 degree from the fovea and abruptly zeroing out in the fovea and optical nerve (Figure 7). Consequently, people experience two blind spots in very dim light

conditions (e.g., under starlight), where the rods play a key role in visual perception (scotopic vision). Signals from dozens to over a hundred rods are converged to one single ganglion cell (Kolb, 2003). This mechanism thereby maximizes the sensitivity of the eye at low light while providing a low-acuity image because the exact source of the signal is ambiguous. The above two factors (namely, high sensitivity and lack of rods) explains why we can see a dim star in peripheral vision, but lose it when staring at it. Light that passes through the photoreceptor layer is absorbed by the pigmented epithelium.

Before entering the photoreceptors, light must go through two other layers, the inner neural layer and ganglion layer. Bipolar cells, horizontal cells, and amacrine cells are three groups of cells in the inner neural layer. Outputs from the photoreceptors pass through the bipolar cells to the ganglion cells. Horizontal cells in the outer plexiform layer provide lateral interconnections between the photoreceptors. Similarly, amacrine cells in the inner plexiform layer link the bipolar cells to the ganglion cells (Ferwerda, 1998). The nerve fibers, axons of the retinal ganglion cells, are packed together at the optical nerve, which is free of photoreceptors and thus commonly known as the blind spot. The optical nerve is located at about 20 degrees away from the fovea on the nasal side and provides the only exit for signals transmitting from the retina to the brain.

3.5 The Axis of the Eye and Related Terms

The non-rotational symmetry of the eye forces researchers to use different terminologies to describe the axes of the eye. Optical axis, visual axis, and fixation axis (Figure 8) are some of the terms frequently found in the visual science literature. These concepts and related terms are important in understanding eye tracking techniques. However, they can

be confusing. In the rest of this section, we provide the distinctive definition of those terms adapted from Atchison & Smith (2000), Carpenter (1988), Cline, Hofstetter et al. (1989), and Grand & Hage (1980).

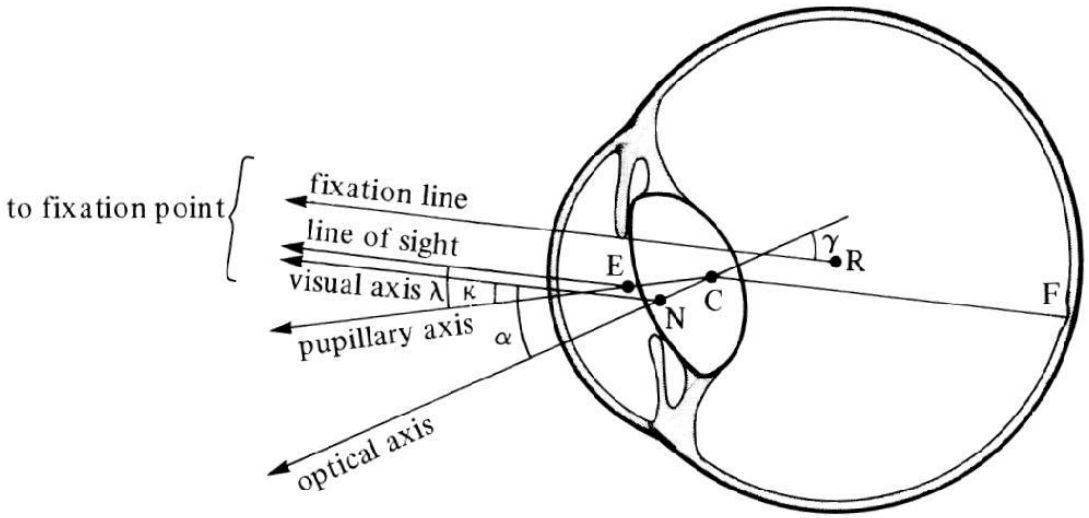

Figure 8. The axes of the eye. From Carpenter, 1988. page 14.

♦ Center of rotation – A point in the eyeball on which the rotation of the eye is centered. This point is not fixed between or even within individuals, and is usually taken as 13 to 15mm from the cornea.

♦ Optical axis – The best approximate line passing through the centers of curvatures of all the primary optical elements in the eye (cornea, pupil and lens).

♦ Visual axis – The line connecting the fixation point, the object nodal point, the image nodal point, and the fovea. It is usually not a straight line as shown in Figure 8 due to the fact that two nodal points are deviated except when the eye looks straight ahead. Since the object nodal point is close to the center of curvature of the cornea, the visual axis is

approximately perpendicular to the cornea. We care most about the visual axis when measuring the direction of gaze.

♦ Pupillary axis – The line perpendicular to the cornea and passing through the center of the entrance pupil.

♦ Fixation axis – The line joining the point of fixation with the rotational center of the eyeball. Ideally this vector should be the one used to describe where the subject is looking. In practice, however, the line of sight is used instead because of the difficulty of measuring the rotational center of the eye.

♦ Line of sight – The line connecting the point of fixation and the center of the entrance pupil. It is also the chief ray of the incoming light that passes through the entrance pupil. In some eye tracking apparatuses (e.g., video-based eye trackers), this vector is used to determine the gaze direction as the visual axis (more specifically, the nodal point) is difficult to measure precisely. If the object is distant, the line of sight is parallel to the visual axis.

- Point of fixation – The point in space that is projected onto the fovea.
- Point of regard (gaze) – A point in space on which the visual attention is directed. It is not necessarily the same as the point of fixation, such as when the eye looks straight ahead while paying attention to a peripherally located object. In the eye tracking context, these two quantities can be considered the same because: (1) it rarely happens that paying attention peripherally occurs while looking straight, and (2) directing the eye to the point where your visual attention focuses is almost instantaneous in normal eye movements (saccades are fast-phase movements, up to 600°/s, see Section 4.2).

The angular relationships between the different axes of the eye shown in Figure 8 are summarized in Table 1. In the image space (retinal surface), the visual axis is on the

temporal side of the optical axis (α = average 5°) and slightly below it (about 2°) (Pedrotti & Pedrotti, 1998). Because of the offset between the line of sight and visual axis, what is measured (the line of sight) in some eye trackers (e.g., video-based eye tracker) does not truly represent the position where the image is projected on the fovea (the visual axis). This discrepancy can be "cancelled out" by a calibration (where the subject looks at known-position points) before the measurement.

Table 1. Angular relationship between the different axes of the eye

	Optical Axis	Visual Axis	Line of Sight
Visual Axis	α	-	-
Fixation Axis	γ	-	-
Pupillary Axis	-	κ	λ

3.6 Paraxial Schematic Eyes

The human eye is a delicate system with some components continuously changing throughout life such that it is difficult to measure the dimension of an individual's eye precisely. Moreover, a model of average eyes is needed so that some studies can be done quickly and efficiently, such as optical calculation and ray tracing. Therefore, a lot of work has been conducted to model the eye with different degrees of accuracy. In this section, Gullstrand Number One Exact Eye and Number Two Simplified Eye for paraxial rays are introduced. The application of the Gullstrand Simplified Eye in calculating the optical relationships of the eye for video-based eye tracking is presented in Chapter 7.

3.6.1 Gullstrand Number One Exact Eye

The Gullstrand Exact Eye is the first optical model incorporating the inhomogeneous nature of the lens (Siedlecki, Kasprzak et al., 1997). Instead of being modeled as two refracting surfaces, the lens is composed of four surfaces, two for the anterior part and the other for the posterior part as shown in Figure 9. The core part of the lens has a higher refractive index than its surrounding edges. The parameters of the model are listed in Table 2.

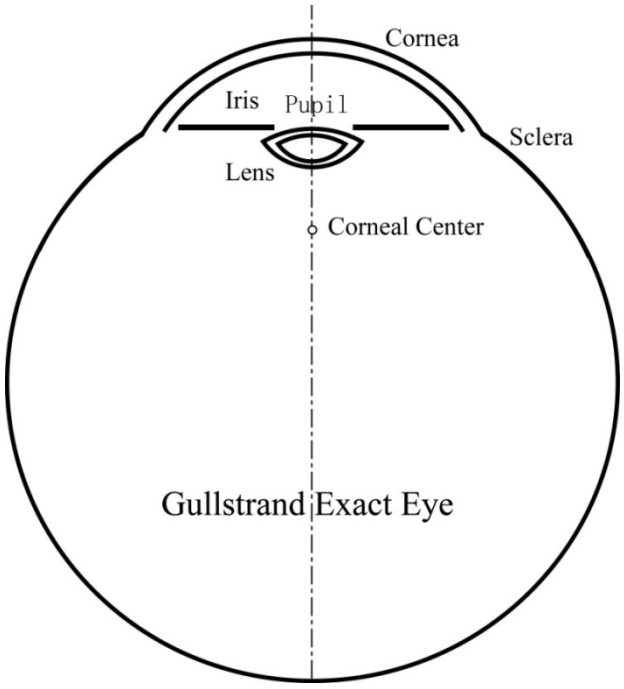

Figure 9. Gullstrand Number One Exact Eye.

Table 2. Gullstrand Number One Exact Eye. The parameters in the parenthesis are for the accommodated eye if they are different from the relaxed eye. Adapted from Grand & Hage, 1980. pp 65-66 and Southall, 1937. pp 57-58.

Optical surface or element	Distance from anterior corneal vertex (mm)	Radius of curvature of surface (mm)	Refractive index	Refractive power(diopters)
Cornea (unit)	-	-	1.3760	43.0530
Anterior surface	0.0000	7.7000	-	48.8300
Posterior surface	0.5000	6.8000		-5.8800
Front principal point	-0.0469	-	-	-
Back principal point	-0.0506	-	-	-
Front focal point	23.2270	-	-	-
Back focal point	-31.0310	-	-	-
Aqueous chamber	-	-	1.3360	-
Lens (unit)	-	-	1.4085	19.11 (33.06)
Anterior surface	3.6 (3.2)	10 (5.35)	1.3860	5 (9.375)
Anterior surface(core)	4.146 (3.8725)	7.911(2.655)	1.4060	2.528 (7.533)
Posterior surface(core)	6.565(6.5275)	-5.76(-2.655)	1.4060	3.472 (7.533)
Posterior surface	7.2000	-6 (-5.33)	1.3860	8.33 (9.375)
Front principal point	5.678 (5.145)	-	-	-
Back principal point	5.808 (5.255)	-	-	-
Front focal point	-69.908 (-40.416)	-	-	-
Back focal point	69.908 (40.416)	-	-	-
Vitreous chamber	-	-	1.3360	-

Total eye (unit)	-	-	-	58.636 (70.57)
Front principal point	1.348 (1.772)	-	-	-
Back principal point	1.602 (2.086)	-	-	-
Front focal point	-15.707 (-12.397)	-	-	-
Back focal point	24.387 (21.016)	-	-	-
Front focal length	-17.055 (-14.169)			
Back focal length	22.785 (18.930)			
Front nodal point	7.078 (6.533)	-	-	-
Back nodal point	7.332 (6.847)	-	-	-
Entrance pupil	3.045 (2.667)	-	-	-
Exit Pupil	3.664 (3.211)	-	-	-
Eye Length	24.3850			

3.6.2 Gullstrand Number Two Simplified Eye

This simplified model consists of three refracting surfaces: a single spherical surface representing the whole cornea, and the anterior and posterior surfaces of the lens (Figure 10). The refractive index of the lens is modeled as homogenous. All refracting surfaces are assumed to be coaxial, with the optical axis of the eye passing through the center of curvature of each element. The pupil is centered at the anterior surface of the lens at the distance of 3.6mm away from the corneal vertex. The refractive index of the aqueous humor is equal to 1.333. Neither this eye model nor the Gullstrand Number One Exact Eye include the eye's rotational center; its distance from the corneal vertex (R=13.5mm) is given in Cornsweet and Crane's paper (Cornsweet & Crane, 1973) describing a

dual-Purkinje eye tracker (a high sampling rate, high accuracy eye tracker that utilizes the 1st and 4th Purkinje images to map the line of sight, see Section 5.4.1). Also, the rotational center of the eye is assumed to be on the optical axis of the eye. The parameters of the eye model are summarized in Table 3 and their symbols used in this dissertation are listed in Table 4. In Chapter 7, this simplified model is used to derive the optical relationships between the pupil and CR (first surface reflection of the cornea) offsets during eye movements.

Table 3. Gullstrand Number Two Simplified Eye. The parameters in parentheses are for the accommodated eye if they are different from the relaxed eye. Adapted from G. Smith and D. Atchison, 1997. pp 778-779.

	Distance from anterior corneal vertex (mm)	Radius of curvature of surface (mm)	Refractive index	Refractive power(diopters)
Cornea	-	7.8000	-	42.7350
Aqueous chamber	-	-	1.3333	-
Lens (unit)	-	-	-	21.755 (32.295)
Anterior surface	3.6 (3.2)	10 (5.0)	1.4160	8.267 (16.533)
Posterior surface	7.2000	-6 (-5.0)	1.4160	13.778 (16.533)
Vitreous chamber	-	-	1.3333	-
Total eye (unit)	-	-	-	60.483 (69.721)
Front principal point	1.550 (1.782)	-	-	-
Back principal point	1.851 (2.128)	-	-	-
Front focal point	-14.983 (-12.561)	-	-	-
Back focal point	23.896 (21.252)	-	-	-
Front focal length	-16.534 (-14.343)			
Back focal length	22.045 (19.124)			
Front nodal point	7.062 (6.562)	-	-	-
Back nodal point	7.363 (6.909)	-	-	-
Entrance pupil	3.052 (2.674)	-	-	-
Exit Pupil	3.687 (3.249)	-	-	-
Eye Length	23.8960			

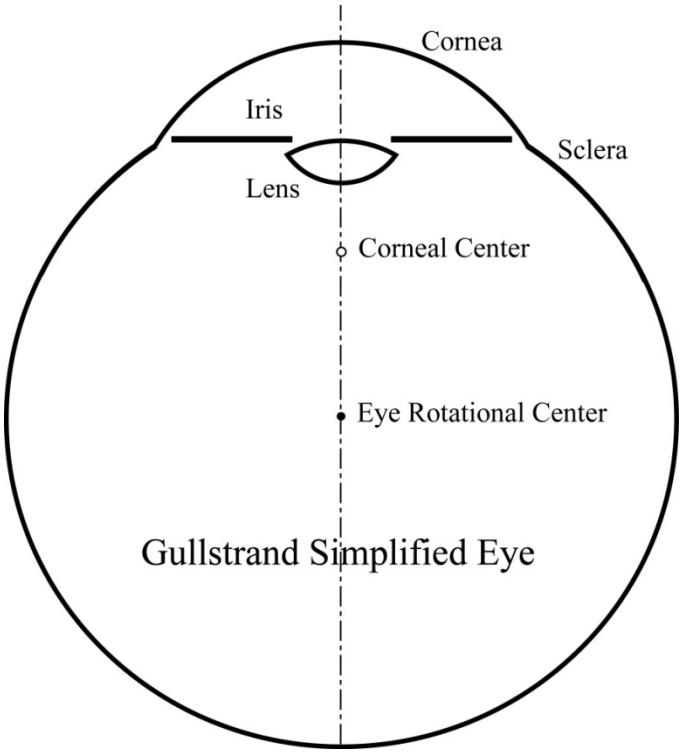

Figure 10. Gullstrand Number Two Simplified Eye.

Table 4. Some symbols used in the Gullstrand Number Two Simplified Eye

	Symbol*
Refractive index of the air	n
Refractive index of the aqueous humor	n'
Radius of the cornea	r
Focal length of the cornea	f
Actual pupil center from the corneal vertex	p
Apparent pupil center from the corneal vertex*	a
Rotational center of the eye from the corneal vertex	R
Illumination Source	S (S′)
Corneal Vertex	A (A′)
CR Position	T (T′)
Focal Point of the Cornea	F (F′)
Corneal Center of Curvature	C (C′)
Eye Rotational Center	E (E′)

*Symbols in parentheses represent the new eye position after a translation or rotation (see Chapter 7).

3.7 Reflections from the Eye

3.7.1 Bright Pupil and Dark Pupil

Bright pupil and dark pupil are two common effects observed when illuminating the eye using an infrared source (Figure 12). The retina is a good retro-reflector; it bounces back

rays in the direction toward the incoming light. If the observer or camera is situated approximately collinear with the illumination source, a bright pupil is seen (Figure 11a). This effect is also well known as “red eye” from flash photographs. On the other hand, the pupil appears as a dark disk compared with its surrounding features if the illumination source is deviated with the line of sight of the observer or the optical axis of the camera (Figure 11b). In order to produce the bright pupil effect, the illumination source needs to be mounted within 3° of the optical axis of the camera (Watanabe, Ando et al., 2003). In other words, a dark pupil is seen if the illuminator is set beyond 3° off the camera’s optical axis.

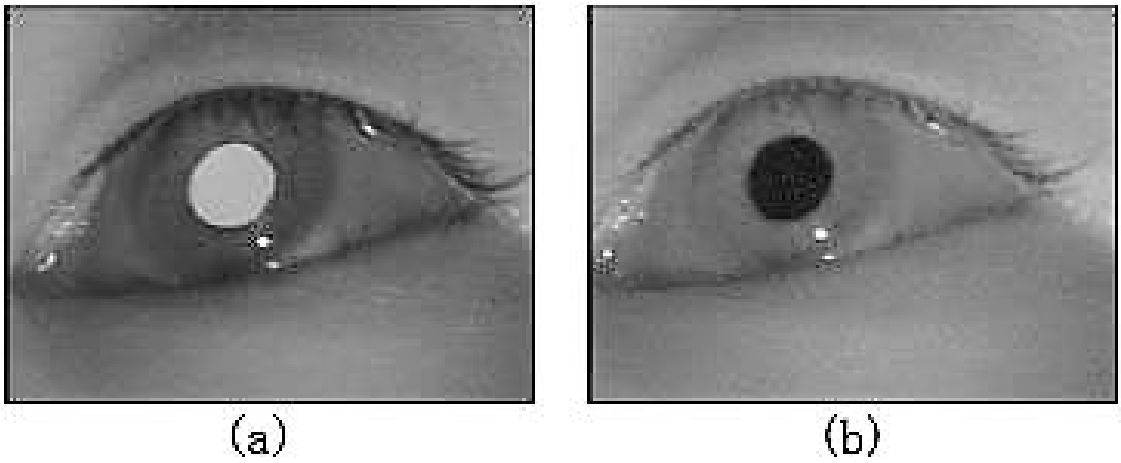

Figure 11. Images of bright pupil (a) and dark pupil (b). Images from C.H. Morimoto, 2000.

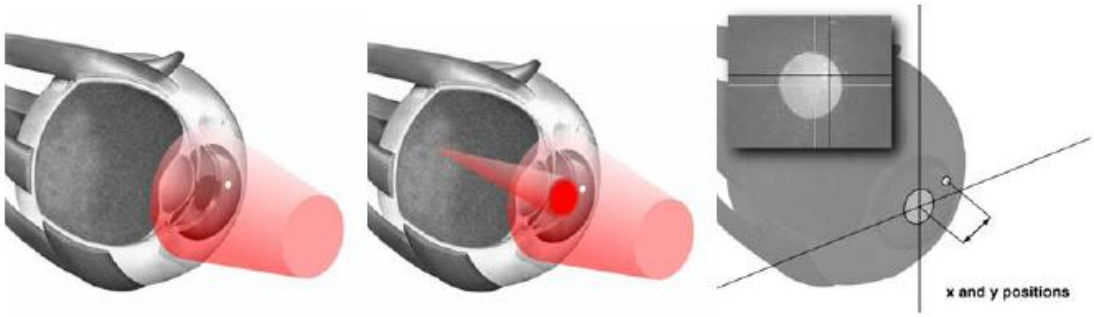

Figure 12. Bright and Dark Pupil. From Jason S. Babcock, MS thesis, Rochester Institute of Technology, 2002.

Both bright and dark pupil effects are widely applied in commercial eye-tracking systems. Applied Science Laboratories 6000 series (ASL, 2005), Eye Response Technologies ERICA (ERT), and LC Technologies EYEGAZE (LCT), etc., mainly rely on the bright pupil, while ISCAN eye tracker (ISCAN) and VISION 2000 (El-Mar), etc., favor the dark pupil.

It is believed that the pupil can be more easily recognized with the bright pupil technique, rather than with the dark pupil technique, when the pupil diameter is above 3.5mm. On the other hand, the pupil can be more detectable with the dark pupil technique when the pupil diameter is below 2.5mm (Borah, 1989). When the pupil diameter is between these two limits, the choice is not intuitive and other factors may need to be taken into account. In general, the bright pupil technique is more robust because it greatly reduces interference caused by eyelashes and other obscuring features, while the dark pupil technique is preferred in complex illumination environments (e.g., outdoors) because it is less sensitive to extraneous infrared sources (Green, 1992).

Hybrid systems, for example, Ebisawa (1995), Haro, Flickner et al. (2000), and Morimoto, Koons et al. (2000), take advantage of both the bright and dark pupil effects.

In these systems, the bright and dark pupil images are stored in the odd and even fields of one video frame, respectively, while the overall illumination of the two images are kept approximately constant. The difference image (high signal-to-noise ratio) between the bright and dark pupils preserves the pupil region while eliminating the background (see Figure 34 C-E).

3.7.2 Purkinje Images

The Purkinje images, specular reflections on the refractive surfaces of the eye, are named after physiologist J. E. Purkinje, who first provided detailed descriptions in 1832 (Davson, 1962, page 108). There are four Purkinje images formed by the light reflected from the front and rear surfaces of the cornea and lens (Figure 13), respectively.

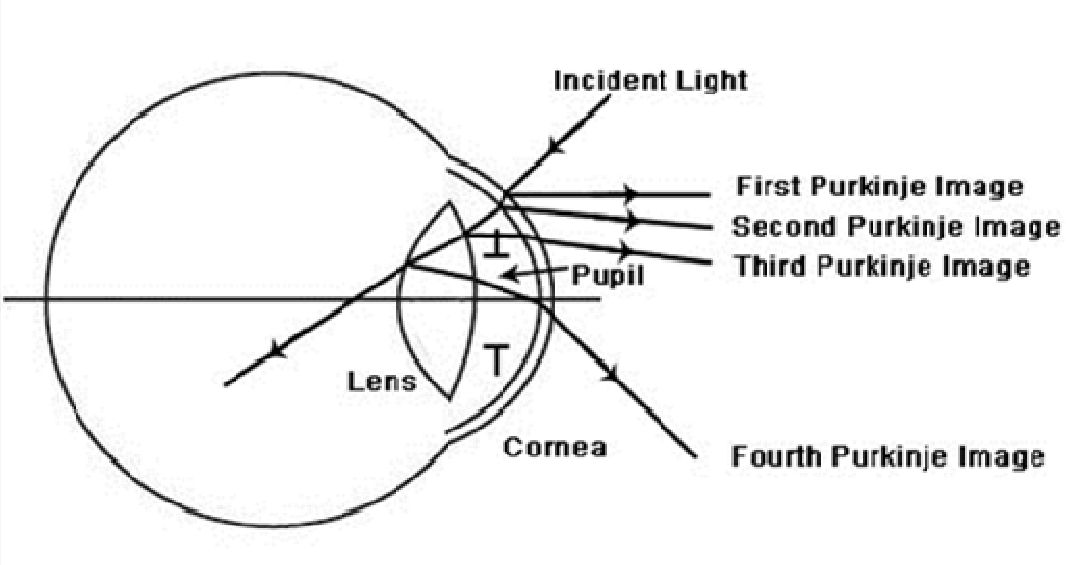

Figure 13. Four Purkinje Images. The four Purkinje images are formed by the light bouncing off the anterior and posterior surfaces of the cornea and the lens, respectively.

The first Purkinje image (P1), also known as the corneal reflection (CR) or glint, is

virtual, erect and much smaller than the object (Figure 14). For an object 20cm high at a distance of 1m from the eye (corneal radius r = 7.7mm), P1 is around 0.78mm high and located approximately at the focal plane of the corneal curvature ($f = r/2$).

The second Purkinje image (P2) is almost coincident with P1, but slightly smaller and much darker (see Table 5 for details).

The third Purkinje image (P3) may be twice as large as P1 (in the relaxed eye), but is usually diffuse because light forming P3 has been partially reflected at the lens capsule (Davson, 1962, page 109). P3 is situated at a plane far away from the others; for the relaxed eye, it is located in the vitreous chamber.

The fourth Purkinje image (P4) is the only real image, which falls on a plane close to P1 and has a similar size (relaxed eye).

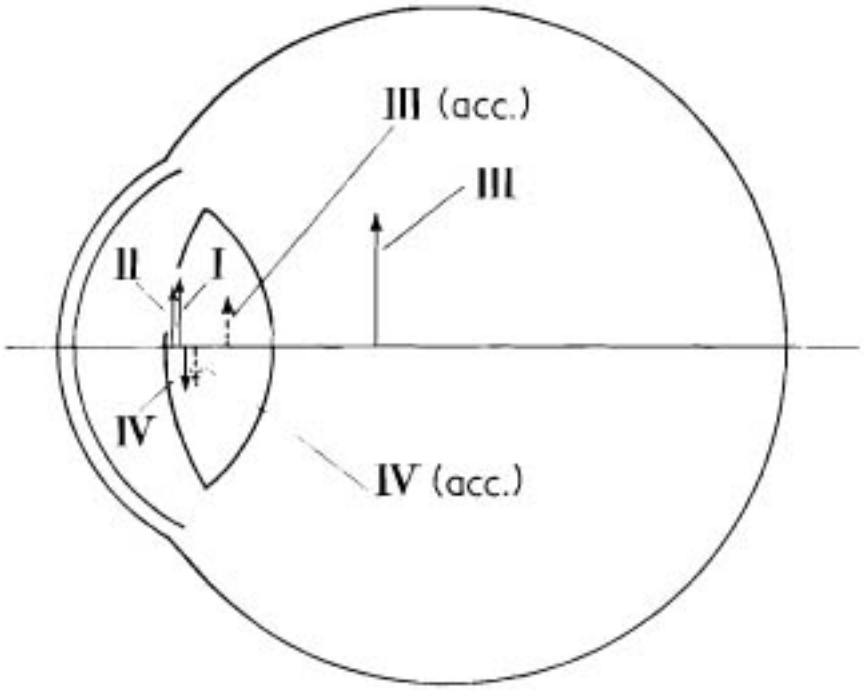

Figure 14. The relative positions and sizes of the four Purkinje images. Image from Hugh Davson, 1962. page 109, Figure 7.

Table 5. Relative positions, sizes and brightness of the four Purkinje images*. Adapted from Davson, 1962. page 111, Table II.

Purkinje Image	Relative Brightness	Relaxed Eye		Accommodated Eye (8.62D)	
		Distance from Corneal Vertex (mm)	Relative Size	Distance from Corneal Vertex (mm)	Relative Size
P1	1.000	3.85	1.00	3.85	1.00
P2	0.010	3.77	0.88	3.77	0.88
P3	0.008	10.59	1.96	5.51	0.74
P4	0.008	3.96	-0.75	4.39	-0.67

*Note that Gullstrand Number One Exact Eye was used in calculating P1 and P2 and Gullstrand Number Two Simplified Eye was used for P3 and P4.

P1 and P4 (Figure 15) can be seen by holding an illumination source to one side of the subject's eye and observing from the other side. P2 and P3 are often hard to see and therefore have not been applied in eye tracking techniques.

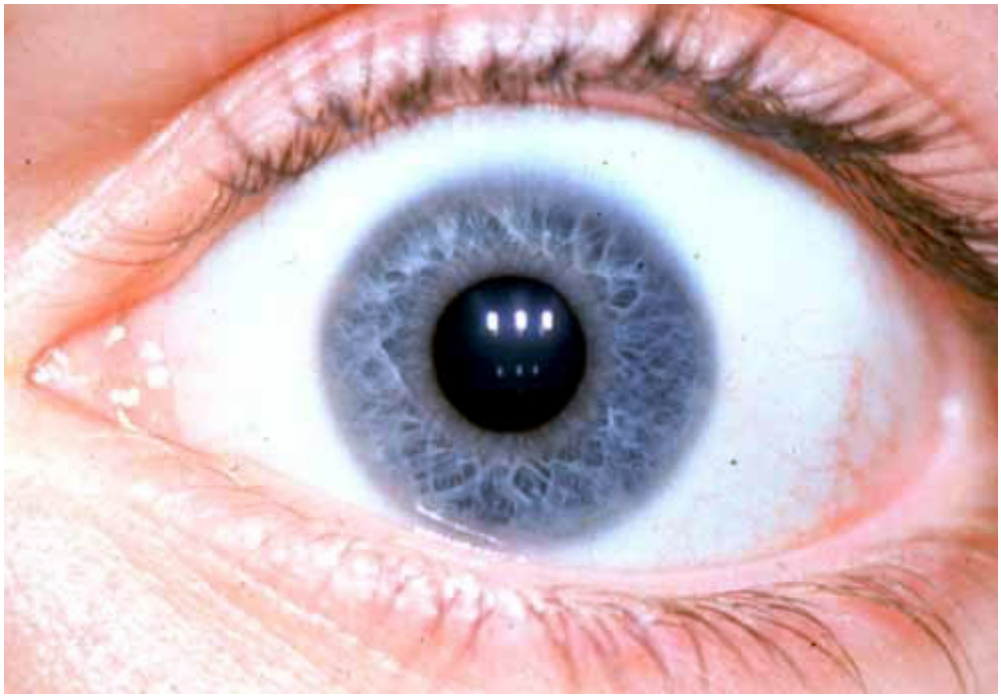

Figure 15. The first and fourth Purkinje images of three illuminators. The brigher images at the top row are P1 and those at the bottom row are P4. Image from http://www.iris-ward.com/_HTM/MEIS/P/1576-MEIS.htm#Top.

The first Purkinje image and the pupil (in video-based eye trackers), and the first and fourth Purkinje images (in dual-Purkinje image eye trackers), are commonly used as combined features in eye tracking in order to reduce the contamination of translational eye movements (relative movements between the eye and imaging sensor due to head movements or camera movements).

3.7.3 Corneal Reflection

A corneal reflection (CR) – also known as the first Purkinje image – is the image of an illumination source formed by the first-surface reflection of the cornea. Because the

radius of the cornea (r = 7.8mm) is much less than that of the sclera (R = 13.5mm), the CR will move in the same direction as the eye rotates. As shown in Figure 16, the cornea serves as a spherical, convex mirror in the image formation process. The CR position can be calculated by using the thin lens equation for paraxial rays (Hecht, 1998, page 159):

$$\frac{1}{s}+\frac{1}{s'}=\frac{1}{f}=\frac{1}{r/2} \tag{3-10}$$

where s and s' are the object and image distances, and $f = r/2$ is the focal length of the corneal surface when seen as a reflector. Note, if the illumination source is moved further away to infinity ($s = \infty$) – a condition in which rays are collimated – its image will lie in the focal plane of the cornea (i.e., the CR is located at $s' = r/2$ from the corneal vertex). Additionally, the CR will be in line with the ray that goes through the corneal center of curvature. These two characteristics of CR movements will be used to derive the important optical relationship between pupil and CR movements in Chapter 7.

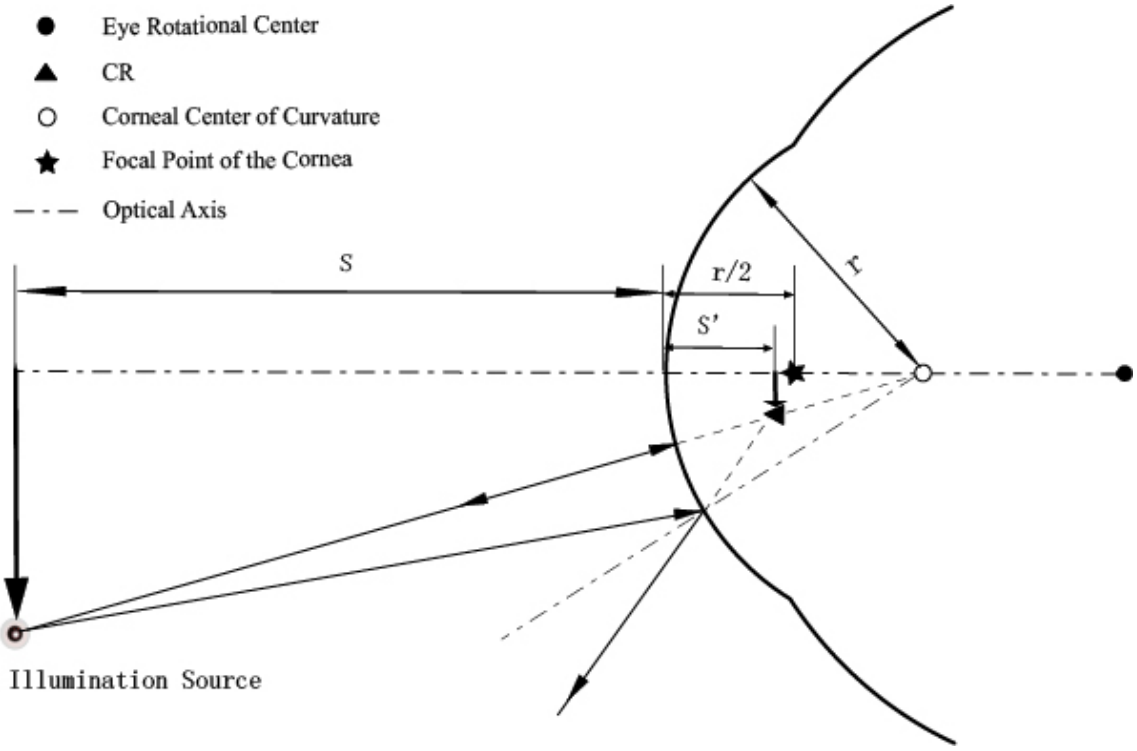

Figure 16. Image of an illumination source formed by the cornea (corneal reflection).

3.7.4 Reflections from Attachments to the Eye

Because the confusion of eye rotations as translations arises from the curved nature of the cornea, by attaching a plane mirror to the eye, an eye tracker can reduce its sensitivity to translational eye movements (Carpenter, 1988, page 418). Such an idea has been applied in several eye movement recording devices, among which the one designed by Alfred L. Yarbus (1967) is undoubtedly very impressive. A series of miniature suction devices, or "caps" (Figure 17), were built and fixed to the subject's eye. The light reflected from the mirror, which is attached to the cap, is collected for measuring eye movements.

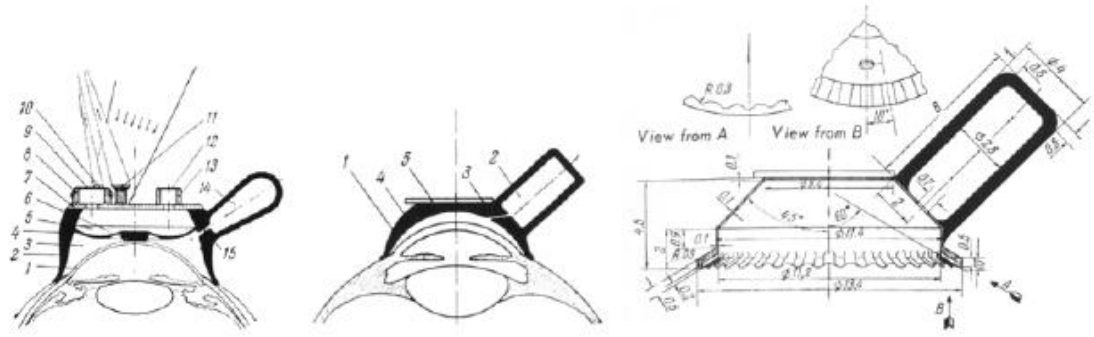

Figure 17. Several suction devices or 'caps' used by Yarbus to measure eye movements. Image from Yarbus, 1967, figure 14, 15 and 17.

Chapter 4 Eye Movements

Due to the limited size of the fovea, which has the highest acuity, the eye can only see a scene about 3~4 square degrees in high detail. We fixate a portion of the scene to gather details by projecting it on the fovea and move our eyes sequentially to gather the detailed fragments from the scene. Eye movements are controlled by three extraocular agonist/antagonist muscle pairs in each eye. The pairs in the ocular system primarily control the eye to move in horizontal, vertical and torsional directions. Generally, eye movements are divided into five categories: smooth pursuit, saccade, optokinetic response, vestibular-ocular reflex and vergence. The pursuit eye movements are of primary interest in the thesis. Its dynamics in a controlled static laboratory and an unconstrained vibrating vehicle are studied in Chapter 9 and Chapter 10, respectively.

4.1 Smooth Pursuit

An airplane flying across the sky appears sharp because you move your eyes to stabilize the image on the fovea, the highest acuity region of the retina. This kind of eye movement is called smooth pursuit. It minimizes or eliminates the relative motion

between the image and the retina (retinal slip). The smooth pursuit system is considered a negative-feedback system (Heinen & Keller, 2004; Pola & Wyatt, 1991) in that the motion is typically triggered by the retinal slip signal. However, a broad range of visual or non-visual information may generate and/or influence the smooth pursuit eye movement, such as non-visual cues (Glenny & Heywood, 1979; Zambarbieri, Schmid et al., 1981), target characteristics (Spering, Kerzel et al., 2005), and knowledge of target motion (Collewijn, Steinman et al., 1985; Kowler, 1989). The smooth pursuit control system can be affected by a number of factors, such as aging (Ross, Olincy et al., 1999), drug abuse (Abel & Hertle, 1988), and schizophrenia (Schlenker, Rudolf et al., 1994), and impairment of smooth pursuit eye movements are diagnostic for neuropsychological dysfunction from Parkinson's disease (White, Saint-Cyr et al., 1983) to dyslexia (Judge, Caravolas et al., 2006).

There are several main parameters used to characterize smooth pursuit (Figure 18). A latency of about 50~150ms (Pola & Wyatt, 1991) is the time delay of pursuit onset in response to a target appearing, which is usually measured by inspecting a velocity trace. The latency can be affected by several factors, for example target velocity and contrast (Ilg, 1997). The interval between the onset of pursuit and reaching a steady-state response is termed the open loop pursuit, during which the eye accelerates to catch up to the target velocity before retinal-slip signals are available for feedback control. During closed loop (steady-state) pursuit, the eye movement has a roughly constant velocity which depends on, and is somewhat lower than, the target velocity. The cumulative error caused by this velocity slip can be compensated by a corrective saccade as shown in the middle panel of Figure 18.

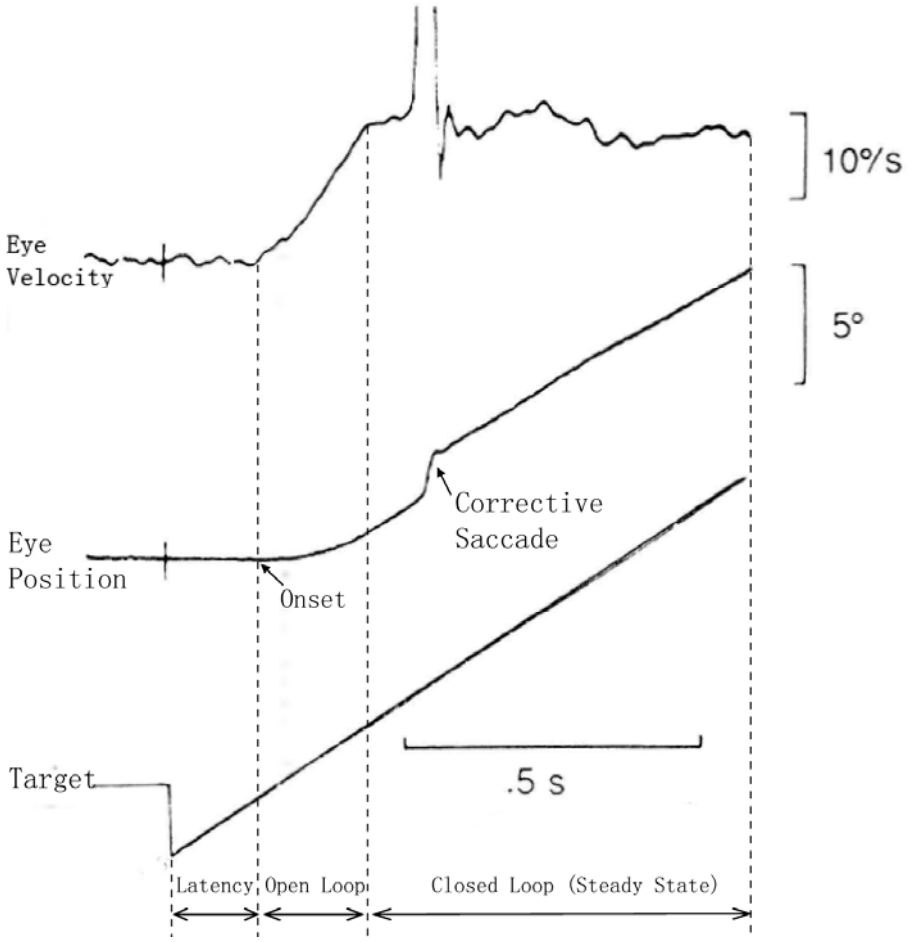

Figure 18. Parameters of smooth pursuit. Task is to track a step-ramp target. Adapted from Pola and Wyatt, 1991.

From the perspective of visual acuity, a smooth pursuit eye movement keeps the image of a target on the region of the retina where visual acuity is high enough for the pursuit task to be performed, minimizing the relative motion between the projected image and the retinal mosaic. Based on the fact that the diameter of the isoacuity area in the retina is approximately 50 minutes of arc (Millodot, 1972), the deficiency of the pursuit system can be tolerated as long as the velocity slip is below a threshold

(accumulated slip limit). A saccadic eye movement (see Section 4.2) brings the image back to this region whenever an accumulated slip exceeds a threshold (Collewijn & Tamminga, 1984). However, the alleviation by a corrective saccade is only temporary, so pursuit eye movements are never completely smooth. Smooth pursuit gain, the ratio between steady-state pursuit velocity and target velocity, is rarely equal to or larger than unity, except for the open-loop period (Pola and Wyatt, 1991).

Though smooth pursuit cannot be initiated voluntarily without a target, it can occur by purposely following a moving target, such as fixating on your fingertip while moving your finger slowly.

4.2 Saccade

Saccades are fast-phase eye movements to bring a target to the fovea, the highest acuity region of the retina. The main parameters of these jump-like movements are characterized in Figure 19.

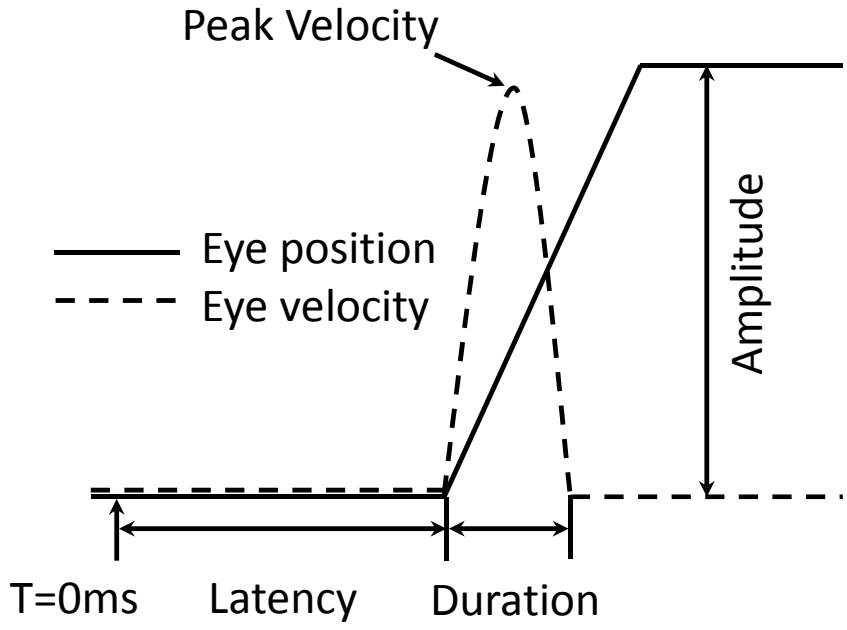

Figure 19. Parameters of saccadic eye movements.

Latency is the reaction time of saccades in response to a target appearing, varying from 100ms to 300ms (Young & Sheena, 1975). Peak velocity is the highest velocity during a saccade, up to 600°/s (Carpenter, 1988). Amplitude is the size of a saccade measured in degrees of visual angle. In daily activities, most saccades are under 20° and large saccades over 30° are often involved with head or body movements (Young & Sheena, 1975). Duration is the time necessary to complete a saccade. The relationship (Figure 20) between the duration and amplitude of saccades can be modeled by the following equation:

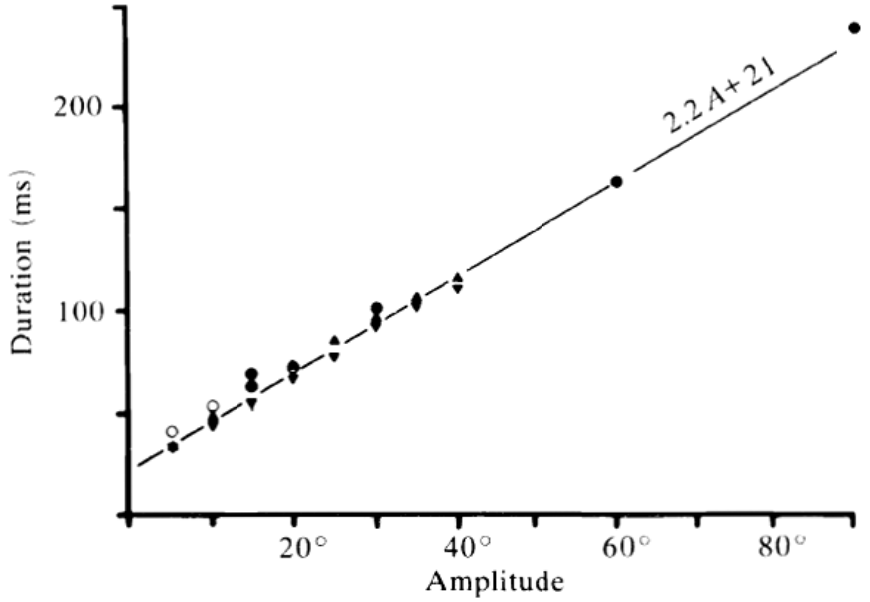

Figure 20. Saccades duration as a function of amplitude. Image from Carpenter, 1988. pp 71, Figure 4.2.

$$Duration = 2.2*Amplitude + 21 \quad (4\text{-}1)$$

where duration is in msec and amplitude is in degrees of visual angle (Carpenter, 1988, page 71).

Saccadic eye movements are stereotyped and ballistic movements. Its approximately fixed relationship among the amplitude, duration and peak velocity is called "The Main Sequence" (the terminology borrowed from astronomical community) (Bahill, Clark et al., 1975), which is independent of the target (Figure 21). During the execution of saccadic eye movements, we do not capture visual information (saccadic masking) because of motion blur.

We are making saccadic eye movements when the eye fixes on one point after another in the visual field.

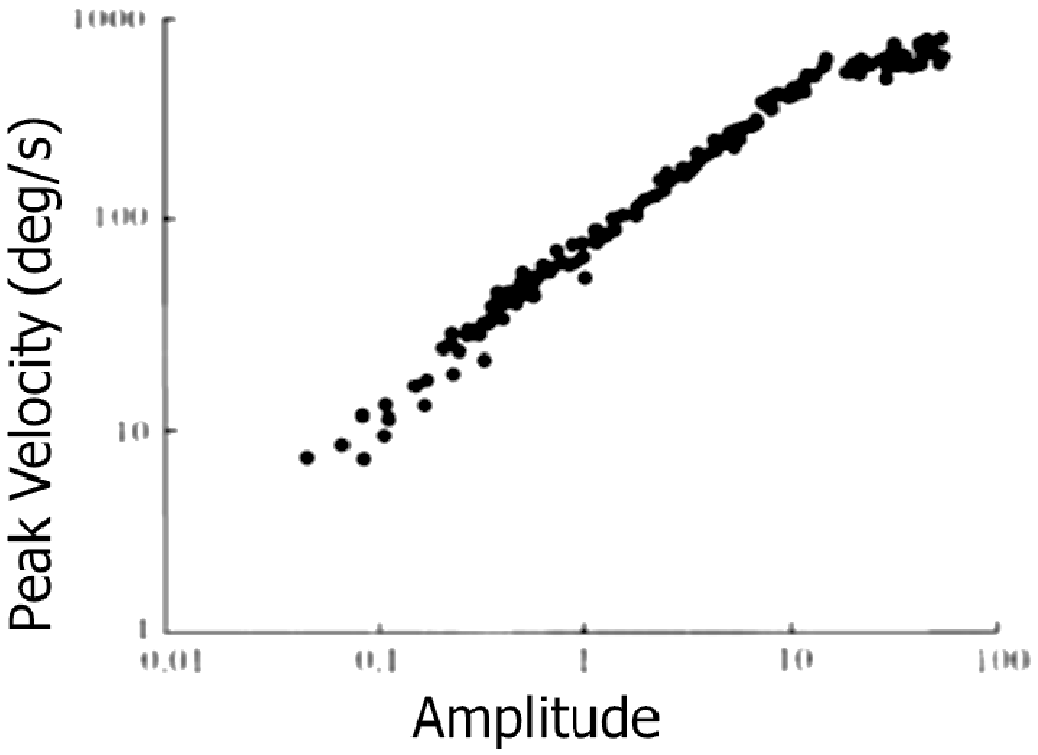

Figure 21. Main sequence of saccades. The relationship between peak velocity and amplitude. Image from Carpenter, 1988. pp 72, Figure 4.3

4.3 Optokinetic Response (OKR)

Optokinetic response (OKR) is an eye movement similar to smooth pursuit, except that it is induced by a large- or full-field stimulus (Collewijn, 1991). Furthermore, OKR and smooth pursuit are often coupled (Carpenter, 1988; Hashiba, Hattori et al., 1996; Pola & Wyatt, 1985; van den Berg & Collewijn, 1986; Yee, Daniels et al., 1983). Some researchers have even stated that the distinction between smooth pursuit and OKR is not clear (Heinen & Keller, 2004). Due to its involvement of large-field motions, OKR often acts in supplementing vestibular-ocular reflex (VOR), the eye movement that serves to

counteract head rotation (see Section 4.4). Optokinetic nystagmus (OKN), describing one specific kind of OKR, is composed of a slow-phase following motion and a fast-phase returning action. Its amplitude varies from 1 to 10 degrees, and it has the maximum frequency of around 5Hz (the maximum frequency of successive returning saccades) (Young & Sheena, 1975).

An example of OKR can be seen when an observer watches a moving train. The eyes first pursue one of the windows and then jump back to catch up the next one.

4.4 Vestibular-Ocular Reflex (VOR)

VOR rotates the eyes to counteract head movements. Three semicircular canals in each inner ear, oriented orthogonally with each other, sense the head movements (Carpenter, 1988).

VOR can be evoked voluntarily, for example in the dark, looking straight ahead and keeping the eye stationary while rotating the head; your eyes remain relatively stationary when you rotate the head because the VOR counteracts the head motion.

4.5 Vergence

During vergence, the two eyes move in opposite directions (disjunction) to align a target on the fovea of each eye. It can be initiated by fixating on your fingertip while moving it closer to, or farther away from, your face; the eyes converge when your finger moves close and diverge when it moves away.

Chapter 5 Eye Tracking Techniques and Applications

The dynamics of eye movements are often studied via eye tracking experiments. Eye trackers provide objective ways to quantitatively measure the subjects' eye movements. Additionally, as only the portion of the scene that is of the most interest and/or importance to the observer are directed by eye movements to be focused on the fovea, the eye tracker techniques serve as objective tools to investigate how an observer gathers information and what attracts a person's interest.

Using different criteria, current eye tracking techniques can be classified by a number of characteristics: (1) bases on measurement results, they can be divided into two types: measuring eye position in the head or measuring that in space; (2) if the obtrusiveness is a key factor, these systems can be sorted as either remote or head-mounted; (3) according to on what physiological features are relied, there are three broad categories: electro-oculography (EOG), scleral search coil, and reflected-light-based techniques.

In this chapter, all the techniques based on the last classification are introduced, while the video-based technique (one kind of reflected-light-based techniques) is covered in detail due to its popularity and relevance to the thesis work. The ideal properties of an

eye tracker and some less popular eye-tracking methods are also included in this chapter. At the end, the applications of eye tracking systems in studying visual and cognitive processes in controlled and unconstrained conditions are addressed.

5.1 Ideal Properties of an Eye Tracker

The following conceptual picture of an ideal eye tracker is mainly based on the experiences of the author and his colleagues with credits to Collewijn (1999) and Hallett (1986). The ideal eye tracker would:

a. Be non-invasive to subjects. Any method that needs an ocular insertion, causes discomfort, and/or blocks the field-of-view of the subject should be avoided. Totally invisible would be a plus. For a head-mounted system, the hardware should be lightweight and not disturb movements of the head;

b. Offer an accuracy and precision of a few seconds of arc;

c. Have high spatial and temporal resolution. Be able to record miniature eye movements on the order of a few seconds of arc with the velocity up to 600 degree/sec;

d. Be capable of measuring three dimensional eye movements: horizontal, vertical and torsional motions. A dynamic range of ±50 degrees for rotational (horizontal and vertical) movements and ±20 degrees for torsional movements should be covered (Collewijn, 1999);

e. Be immune to translational eye movements (due to headband slippage, muscle tremor, camera movement, etc.);

f. Have no crosstalk between different measurements. For example, horizontal signals of eye movements should not be coupled with vertical and torsional ones;

g. Allow unrestricted physical motions (eye, head and body) and easily be integrated with a head or body tracker;

h. Be applicable to a large population in a wide range of tasks;

i. Ensure robustness in different working conditions, resisting performance degradation caused by ambient illumination, heat, magnetic field and other factors in the environment;

j. Ascertain easy set-up, calibration and operation;

k. Permit extendibility to binocular measurements;

l. Exhibit a good linear correlation between eye movements and output data, or ensure that the nonlinearity can be accounted for by a simple relationship. Any severe nonlinearity is troublesome and hard to be corrected completely by a calibration;

m. Operate in real time with a negligible time lag.

If this ideal eye tracker would exist, it would be applied to any eye tracking studies with a satisfying performance. However, eye trackers today, no matter on what technique they are based, fall short on some of the above ideal properties, and all these weaknesses are unlikely to be conquered completely in the foreseeable future. In the following sections of this chapter, the primary techniques of eye tracking are introduced and their performances in comparison with the ideal eye tracker are discussed.

5.2 Electro-Oculography (EOG)

The electro-oculography technique relies on the measure of corneo-retinal potential differences when the eye rotates by placing electrodes at the outside canthi of the eye (Figure 22). By reason of the higher metabolic rate at the retina, the potential in the cornea remains positive. The recorded potentials at the surface are in the range of 15 to 200μV, with nominal sensitivities on the order of 20μV/degree. EOG measurements are analog and therefore have a very high temporal resolution (high bandwidth). This method provides a large measurement range of up to ±70 degrees but with a typical accuracy of only ±1.5~2 degrees (Young & Sheena, 1975).

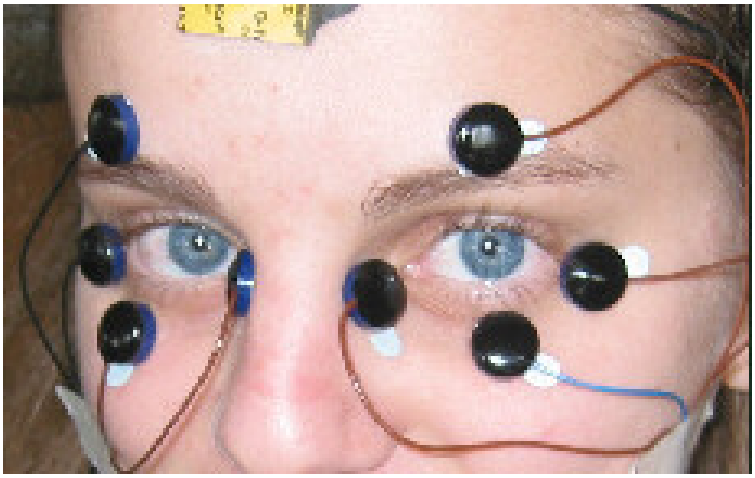

Figure 22. A subject wearing an EOG apparatus. Image from Metro Vision, www.metrovision.fr.

The slow drift of EOG signals and analog outputs make it a good option to measure high frequency components of eye movements, such as saccades, and velocity and acceleration profiles of eye movements (Borah, 1989). Furthermore, because it doesn't require the visibility of the eye, EOG may be preferable in certain applications, for

instance, recording eye movements during sleep.

The downsides of EOG include muscle-action artifacts, nonlinearity beyond 30 degree of visual angle, and signal drifting due to metabolic activities (diurnal variation, alertness state, light adaptation, etc.) (Aslin & McMurray, 2004; Young & Sheena, 1975). Some issues can be overcome to some degree by conducting multiple calibrations during measurements, however, the calibration of an EOG system itself turns out to be non-trivial.

In comparison with an ideal eye tracker, EOG fulfills (d) (excluding torsional eye movements), (h) and (k), but not (a), (b), (f), (i), (j) and (e).

5.3 Scleral Search Coils

This technique evolved from an optical lever method, which collects light bouncing off a flat mirror attached to the eye (see Section 3.7.4). The scleral search coil technique requires insertion of two wire coils, embedded in a contact lens, onto the sclera or limbus (the interface between iris and sclera). The coils induce voltages when they perturb a large electromagnetic field surrounding the subject's head (Figure 23). The voltage is read out through a thin, flexible wire stretching from the lens. In order to fit tightly to the eye to avoid slippages, a negative pressure under the contact lens is usually introduced by dripping a special solution or extracting a small amount of fluid from the eye.

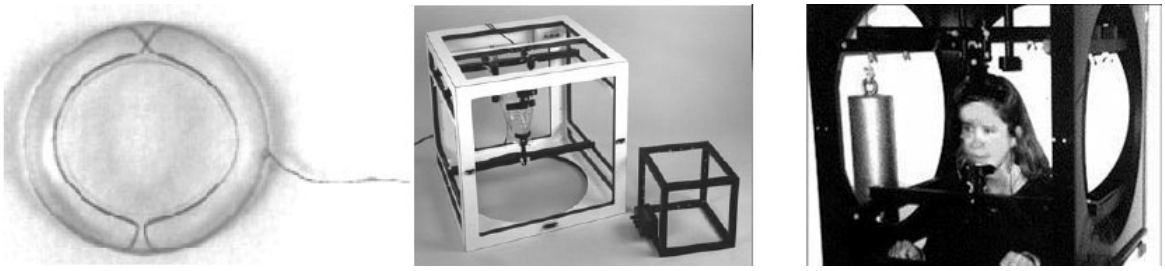

Figure 23. Scleral search coil technique. Left: coils embedded in a contract lens. Middle: frame setting surround the subject's head which produces a large electromagnetic field. Right: whole system setup. Images from Skalar Medical BV.

This technique is one of the most accurate methods in eye tracking and widely regarded as "gold standard" (Collewijn, 1999; Irving, Zacher et al., 2003). Its accuracy is about 5~10 arc seconds within 5 degrees (Young & Sheena, 1975). It can also be extended to three-dimensional and real-time measurements without any restriction of visual field and lighting condition. This technique even works when the eyes are closed.

However, with the high accuracy comes a limited measurement range. Therefore, the scleral search coil technique is not recommended for measuring eye movements larger than 5 degrees. Due to the difficulty in the insertion of contact lens (Figure 24), great care by an experienced operator is necessary. This method may also cause ocular discomfort, slight irritation of the conjunctiva, increased intraocular pressure, and degradation of visual acuity (Irving, Zacher, et al., 2003). Therefore, twenty minutes is usually the maximum duration for wearing the scleral search coil.

Figure 24. Insertion of the scleral search coil. Images from Skalar Medical BV.

In addition, the scleral search coil is found to limit the dynamics of saccades. Saccades last about 8% longer and are approximately 5% slower when both of the eyes wear coils (Frens & van der Geest, 2002). Most researchers agree that all methods requiring insertion of components to the eye (scleral search coils, optical lever, etc.) are unsuitable for human subjects, whenever natural motions are involved.

This gold standard fits well the ideal criteria in (b), (c), (d), (i) and (k), while it is less effective in (a), (g), (h), (i) and (j).

5.4 Non-Image-Recoding Based Eye Tracking

This is a broad concept including a wide variety of techniques, but they are commonly based on collecting light reflected from variant surfaces of the eye.

A severe problem in these techniques is that eye rotations are usually coupled with eye translations (see Section 7.1 for detail). Translational eye movements, as a result of sensor movement, headband slippage or muscle tremor, can contaminate rotational eye movements and thus introduce measurement errors. Every 0.1 mm translational eye movement introduces approximately one degree of error (Carpenter, 1988; Crane. &

Steele, 1985; Young & Sheena, 1975). This kind of translational confusion can only be eliminated completely by attaching supporting components into the tissues of the bony skull (Collewijn, 1999).

Today there are two common approaches to minimize the coupling effect of eye translations: one is to record both the first and fourth Purkinje images (dual Purkinje image eye trackers), and the other is to calculate the positional difference between the pupil center and CR (video-based eye trackers, see Section 5.5).

5.4.1 Dual Purkinje Image Eye Tracker (DPI)

The DPI tracker depends on tracking the positions of the first (P1) and fourth Purkinje (P4) images. It is categorized as a non-image-recording based technique because no images are recorded by discrete (pixel-based) sensor; continuous sensors are used instead. In the DPI tracker, the contamination from eye translations in eye rotations is minimal because the positional difference vector between P1 and P4 is only sensitive to eye rotations. The functional relationship (Cornsweet & Crane, 1973) between the difference vector (S) and eye rotational angle (θ) is described by:

$$S \approx 7 \sin\theta \qquad (5\text{-}1)$$

where S is in millimeters and θ is in degrees.

The first and fourth Purkinje images are projected optically onto four-quadrant photo detectors, which yield analog signals proportional to the off-center displacements of P1 and P4. The two Purkinje images are superimposed to be concentric by adjusting a series of mirrors for each image via servomotors. The final outputs are electrical signals, which are generated by the servomechanism to maintain center superimpositions.

Starting in 1965, DPI eye trackers have evolved through several generations of developments by SRI (Stanford Research Institute) International. In the newest generation 6 (Figure 25) by Fourward Optical Technologies, Inc. (a company authorized by SRI International to market the DPI tracker), it has an accuracy and resolution on the order of one minute of arc or smaller. The response time is about 1ms and the slew rate is greater than 2000 degree/sec (limited by the servomotors driving the mirrors), which is high enough even for the fastest eye movement (Fourward). By attaching an extra component, the system can also be extended to track accommodation changes.

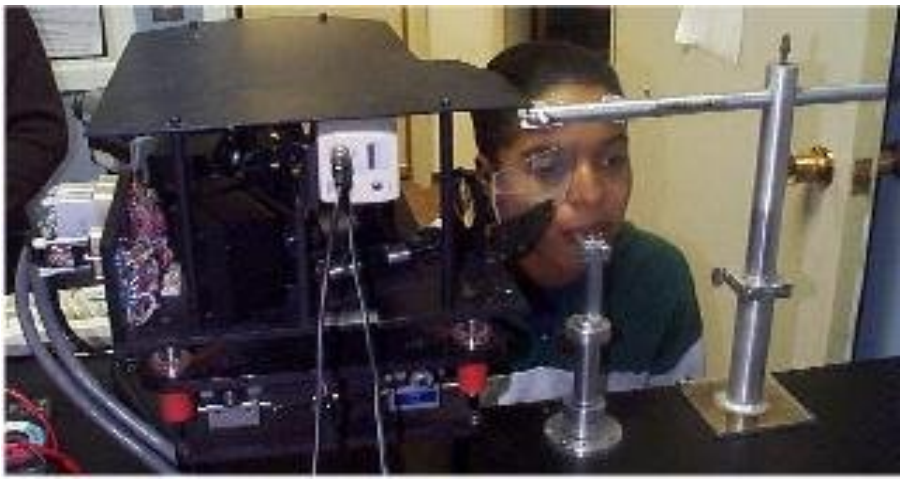

Figure 25. The dual Purkinje image eye tracker. Image from Fourward Technologies, Inc.

The main disadvantages of this technique are the limited operational range (up to ± 20 degrees), the necessity of strict stabilization of head motions (use of a bite bar and/or head rest), and the requirement of a relative high-intensity illumination source. The last requirement as well as its applicable usage in a controlled lighting condition is essential to obtain an appropriate fourth Purkinje image to track. The measurement range can be extended up to ± 25 degrees by dilating the pupil. (P4 is easily masked by the iris in a

wide oblique light.) Moreover, due to the lag of the lens of the eye at the beginning of saccades and overshoot at the end, the peak velocity of the saccadic movements measured by the DPI tracker is substantially greater than that by the scleral search coil (Deubel & Bridgeman, 1995a, 1995b), as shown in Figure 26.

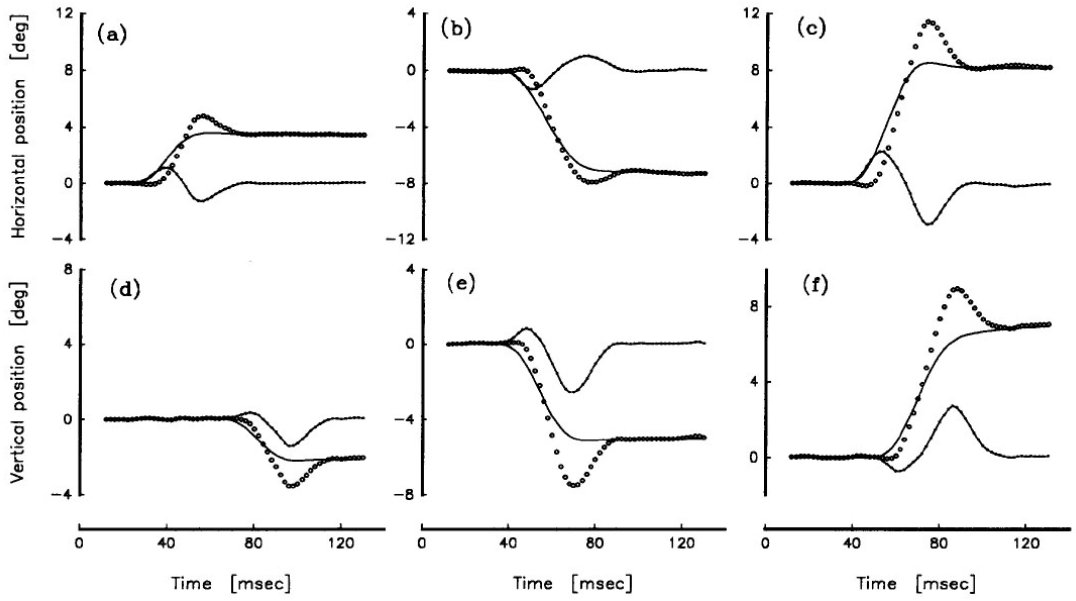

Figure 26. Saccades of different sizes recorded simultaneously by the scleral search coil (solid lines) and the Purkinje eye tracker (solid dots) along with their difference vectors (solid lines close to the baseline). Image from Deubel, et al, 1995.

From the perspective of an ideal eye tracker, a DPI tracker satisfies (a), (b), (c), (e), (f), (k) and (l), but is limited in (d), (g), (i) and (j).

5.4.2 Limbus Eye Tracker

The limbus, the sharp boundary between the iris and sclera, is an easily identifiable edge

which can be used to assess the line of sight of the observer. In a limbus eye tracker, photodiodes are mounted on a spectacle frame to record the light reflected from the eye (primarily from the sclera) (Figure 27). When only horizontal eye movements are of concern, the limbus position, from which the eye position is derived, can be determined by measuring the difference of the reflected light from the fixed areas on each side of the line of sight. When vertical eye movements also need to be measured, due to the high possibility of occlusion of the iris by the eyelid, other information, such as pupil position, eyelid level, or vertical motions of a visible part of the limbus, are necessary to determine the eye position (Young and Sheena, 1975).

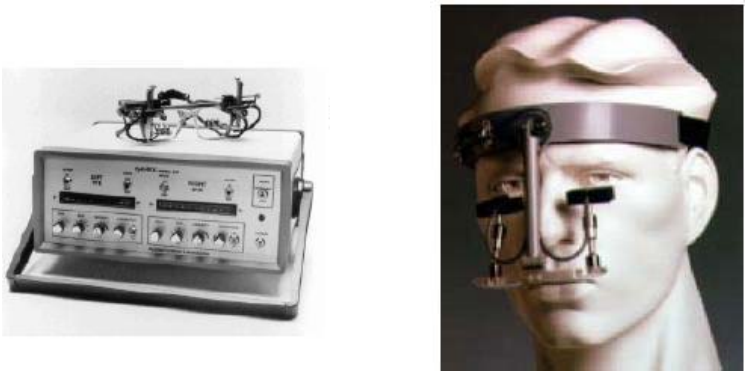

Figure 27. Limbus eye trackers. Left image from Applied Science Laboratories (www.a-s-l.com), right image from Microguide, Inc. (www.eyemove.com)

The limbus eye tracker is easy to set up and not substantially obtrusive to the subject. Its main problems include the confusion of translational eye movements, the difficulty in the measure of vertical eye movements, and the severe crosstalk when measuring horizontal or vertical eye movements.

The limbus eye tracker scores well with the ideal eye tracker in (a), (g) and (k), but falls away in (b), (c), (d), (e), (f) and (i).

5.5 Video-Based Eye Tracking

Most eye tracking techniques introduced so far have unfavorable issues such as great invasiveness to the subject (e.g., scleral search coil), contamination with head movement (e.g., limbus eye tracker), or requirement of strict head stabilization (e.g., DPI eye tracker). Taking advantage of rapid developments in optical sensors (e.g., CCD and CMOS) and image processing techniques, video-based eye trackers conquer, to some extent, these problems and have become one of the most widely used eye tracking techniques.

5.5.1 Fundamentals

Video-based eye tracking systems use a video camera to image the eye. Based on how this "eye camera" is positioned, the systems can be divided into two categories: head-mounted and remote. Head-mounted eye trackers typically include an additional camera to image the scene at which the subject is looking while remote trackers commonly track the observer's gaze on a computer monitor on which the subject is performing a task. Head-mounted video-based eye trackers include an optical module fixed to a headgear, which is then worn by the subject. They are more intrusive than remote trackers but provide more freedom of motion to the subject. The recent rapid improvements in compact electrical and optical components have greatly reduced the size of head-mounted modules (e.g., Babcock & Pelz, 2004; Li & Parkhurst, 2006). On the other hand, remote video-based eye trackers have the eye camera mounted remotely on another base, which can be completely unobtrusive at the sacrifice of overall flexibility.

Typical video-based eye tracking systems simultaneously record an image of the pupil and the corneal reflection (CR) of an illumination source. Although the pupil and CR as seen in an eye image are both sensitive to translational eye movements with respect to the eye camera (caused by headgear slippages, sensor movements, muscular tremors, etc.), their *positional difference vector* (P-CR) provides a signal that changes primarily with rotational eye movements in the head (Merchant, Morrissette et al., 1974) (Figure 28). As such, this video-based technique is moderately tolerant of translational movements of the eye with respect to the sensor, which may be a serious problem in some other reflected-light-based techniques (e.g., limbus eye trackers). The relationship between the P-CR vector and eye rotational angle is represented in the following equation with an accompanying diagram in Figure 29.

$$P-CR = K\sin\theta \tag{5-2}$$

where $K \approx 4.75mm$ is the distance between the apparent pupil center and corneal center, and θ is the angular gaze direction. Note that in this equation the incoming rays are assumed to be collimated (see Figure 29).

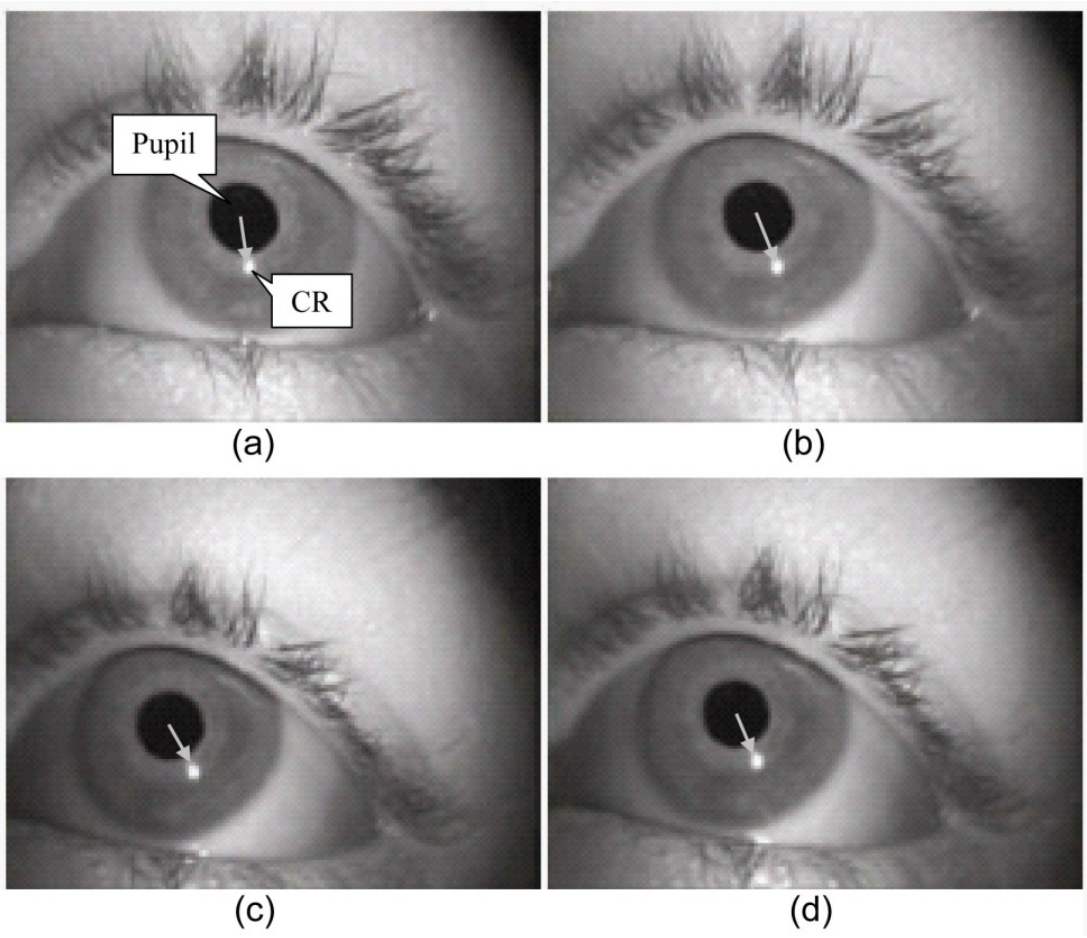

Figure 28. Eye images before (a) and after (b) a rotational eye movement; the P-CR vector has noticeably changed. Eye images before (c) and after (d) an apparent translational eye movement caused by a camera movement; the P-CR vector has changed only slightly. Images from Kolakowski & Pelz (2006).

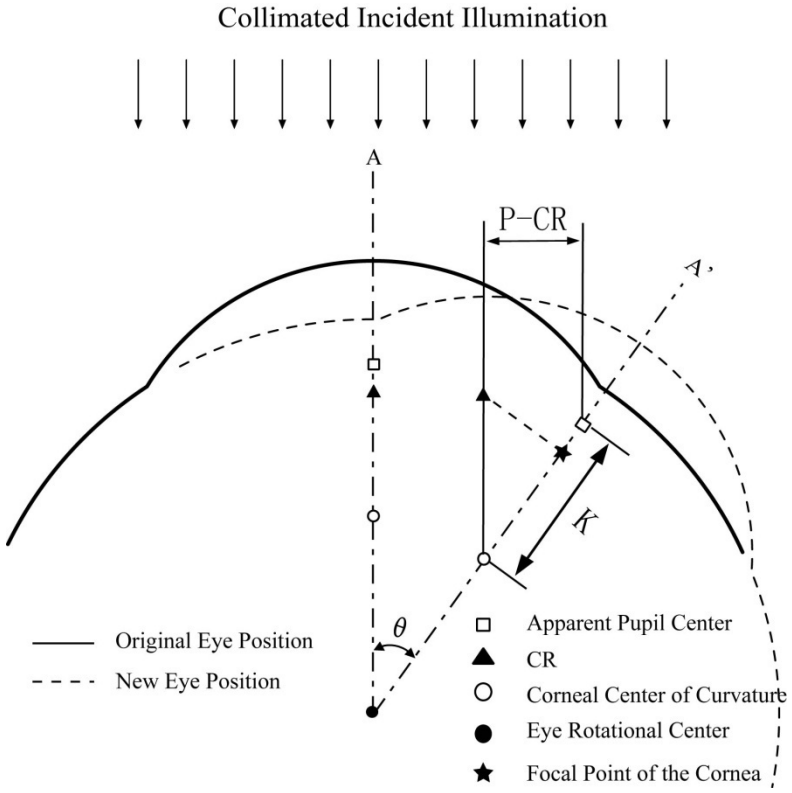

Figure 29. The relationship between the P-CR vector and eye rotational angle. Note that the CR always falls on the focal plane of curvature of the cornea.

Near infrared (IR) illumination is commonly used in video-based eye tracking. IR provides the benefit of being invisible to humans so it does not disturb or distract the observer. Additionally, the iris reflects strongly in the near IR, irrespective of iris color. The strong IR reflectance yields high-contrast images that are particularly beneficial to pupil detection. An infrared-sensitive camera with a filter blocking the visible spectrum is usually deployed to improve the specificity of the detection.

5.5.2 Illumination Structure

Although the video-based eye tracking technique has not been revolutionarily improved since it was innovated in the 1770s (Merchant, Morrissette, et al., 1974), there are a number of different illumination structures used in the laboratory and commercial eye trackers to optimize performance.

Eye-Trac 6000 by ASL

The optical module in the Eye-Trac 6000 by Applied Science Laboratory, Inc. (ASL) is available in several different configurations and a standard version is shown in Figure 30. A lens after the infrared LED (IRED) is used to produce collimated rays shining on the eye (collimated illumination).

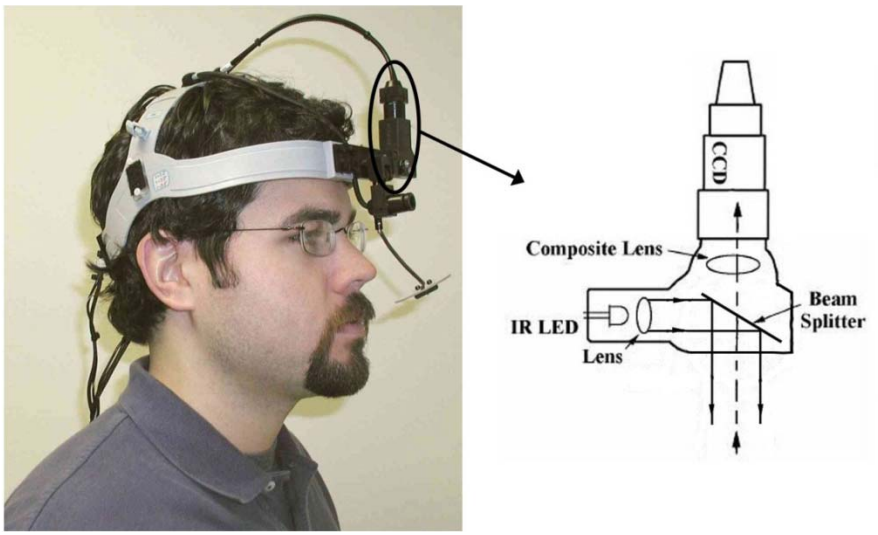

Figure 30. ASL Eye-Trac 6000 eye tracker. The infrared illumination from the LED is relayed by the visor to shine on the eye. The light reflected by the eye is collected by a solid state sensor (CCD) in the head-mounted module. The visor is coated to reflect the infrared illumination and transmit the visible spectrum. Left image from Applied Science Laboratory, Inc.

To eliminate disturbing visible light, an infrared passing filter is set in front of the eye camera, which is chosen to be sensitive to infrared light, as well. A visor (hot mirror) slightly below the eye is used to relay rays to illuminate the eye. Additionally, this configuration ensures that the field of view of the subject is not blocked. The scene camera is mounted on the headband to capture what the subject is seeing. The camera can also be mounted vertically under the visor so that it can achieve the optical equivalent of placing the scene camera in the eyeball (ASL, 2005), eliminating parallax.

Vision 2000 by EI-MAR

The system has a wide measurement range, horizontally ±40° and vertically ±30°. The extension is achieved by using two infrared sources to keep at least one CR trackable even when the CR rolls over the scleral region (Figure 31) (DiScenna, Das et al., 1995; Irving, Zacher, et al., 2003). Another advantage of using multiple sources is that the eye images are easy to be illuminated evenly. Unfortunately, no technical detail about the system has been made available publicly.

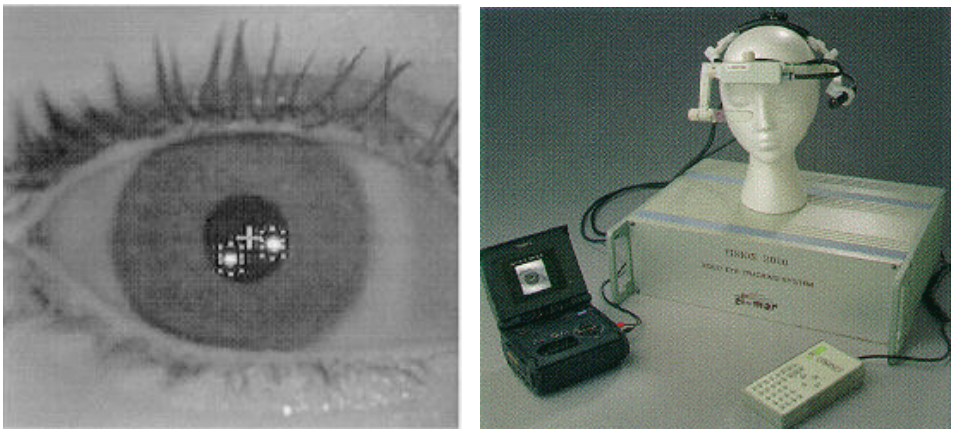

Figure 31. Vision 2000 by EL-MAR Inc. Images from EL-MAR Inc.

Eyelink II by SR Research

The system uses three discrete infrared LEDs to illuminate each eye (Figure 32). Two of them are mounted at an equal distance from the eye camera in order to have an even illumination in the eye, while the other is set close to the head camera on the frame for generating the CR. When working in the P-CR mode, the two "pupil LEDs" and one "CR LED" flash alternatively at the maximum rate of 250Hz. When recording the pupil only, the sampling rate can be increased up to 500 Hz (SR_Research, 2002). SR Research Ltd. also offers a high-speed video-based eye tracker (EyeLink 1000), which is a head-supported system capable of 1kHz sampling rate. Both eye trackers use custom high-speed eye cameras.

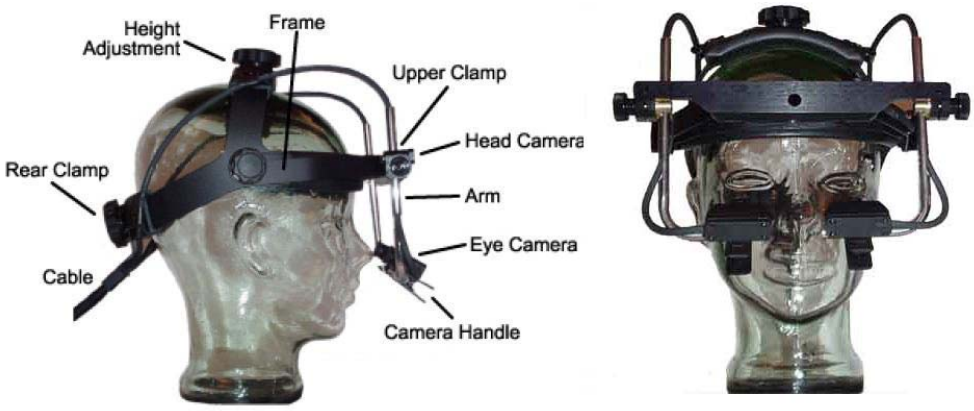

Figure 32. Eyelink II head-mounted system. Images from Eyelink II user manual. version 1.05, 2002.

ISCAN Eye Tracker

The system is based on the dark pupil effect (Section 3.7.1), as the IRED is set off the optical axis of the eye camera (Figure 33). The gaze estimation is computed by either an online or offline processing (ISCAN, 1996).

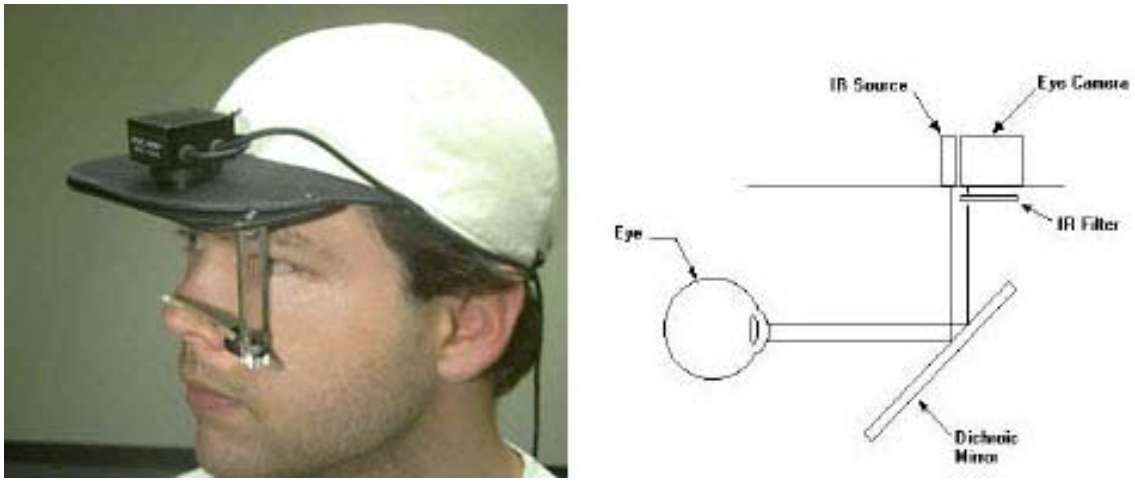

Figure 33. ISCAN eye tracker. Left image from of ISCAN Inc.

IBM BlueEyes

This system consists of two sets of LEDs (Figure 34). One set is mounted very close to the optical axis of the camera, producing a bright-pupil image. The other is relatively far off the optical axis so as to generate a dark-pupil image, while maintaining approximately the same overall illumination as does the on-axis set. These two sets of LEDs are alternatively turned on in synchrony with the two fields of an interlaced video frame. A low noise pupil image is obtained by subtracting the dark pupil image from the bright pupil image (Morimoto, Koons et al., 1999). A pan/tilt base (Figure 34B) is added to search for the eye at the beginning or reinstate the tracking once losing it. Although the system works adequately in real-time, it loses spatial and temporal resolutions due to the de-interlacing of the video frames in order to extract bright and dark pupil images.

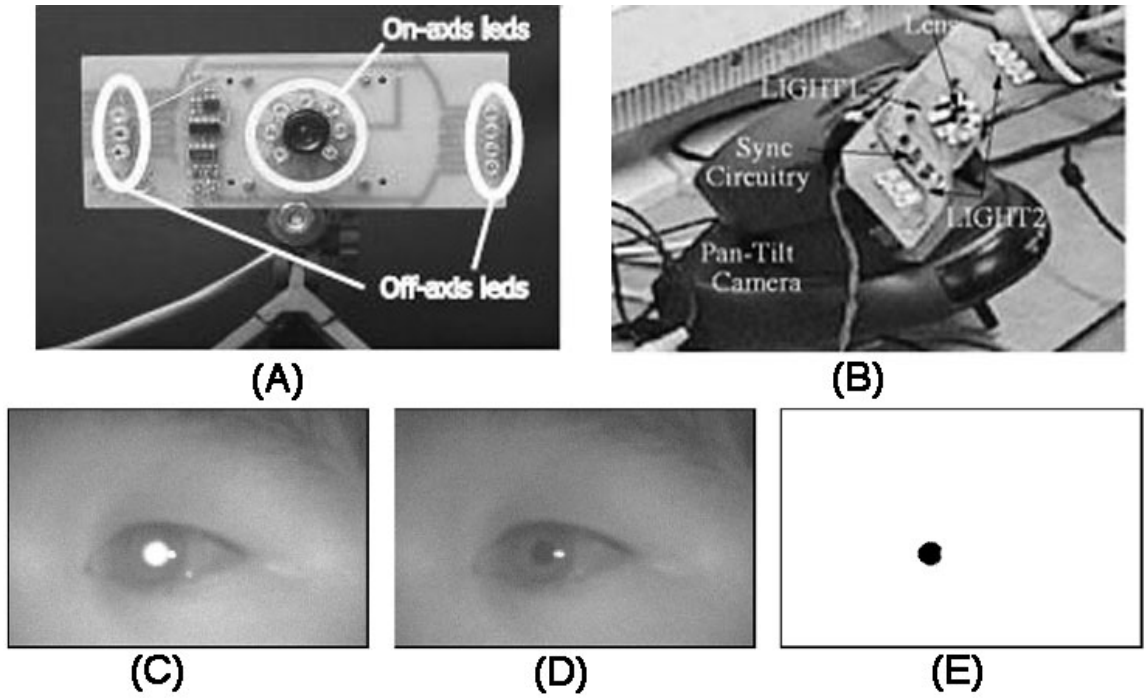

Figure 34. IBM BlueEyes. (A) Camera and two sets of LEDs. (B) Pan/tilt base. (C) Bright pupil image. (D) Dark pupil image. (E) Difference image after thresholding. Images from C.H. Morimoto, 1999.

RIT Lightweight Eye Tracker

This tracker has a lightweight headgear with a simple optical configuration (Figure 35). An infrared LED, mounted about 30~40mm away from the eye, is mounted to the side of the eye camera so that dark pupil images are produced (Babcock & Pelz, 2004). The arrangement, by setting the light source and miniature eye camera on an extended arm that point to the eye, however, sacrifices a small portion of the field of view of the subject.

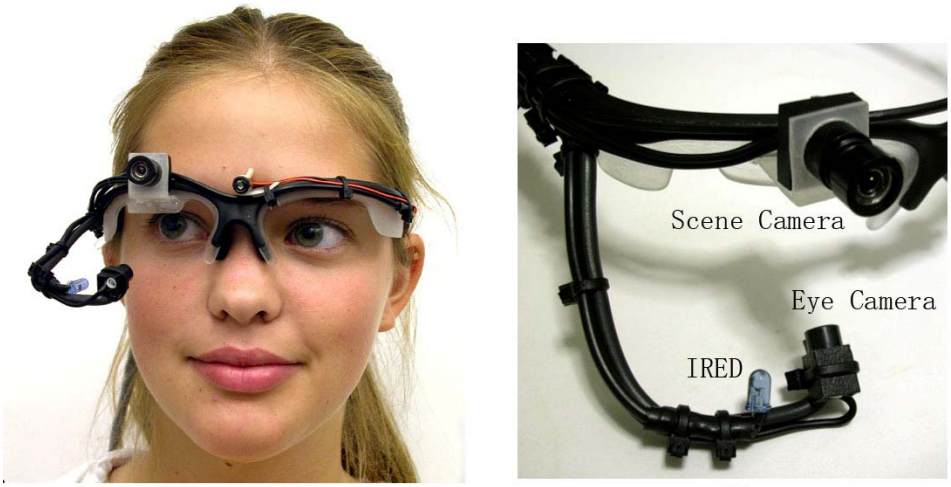

Figure 35. RIT lightweight eye tracking headgear. An infrared LED is mounted beside the optical axis of the eye camera so that it produces dark pupil images. Images from Babcock and Pelz, 2004.

Although the video-based method provides a promising way to measure eye movements, it inevitably has weaknesses in many aspects. Because one important piece of this thesis work is to optimize a head-mounted video-based eye tracking system, a separate section (Section 7.1) is included to address those limitations.

The video-based eye tracker matches the ideal one moderately in (a), (d), (e), (f), (g), (j), (k) and (l) and falls short of ideal in (b), (c) and (i).

5.6 Hybrid Systems

As mentioned earlier, all currently available eye tracking techniques have inherent weaknesses; some researchers are thinking of combining different techniques to overcome the limitations by an individual method. Triesh et al. (Triesch, Sullivan et al.,

2002) proposed such a system that integrated two kinds of eye trackers into a single head-mounted system. A limbus tracker (ASL 210) sampling at 1.25kHz was introduced to detect the onset of saccades with low latency, while a video-based tracker (ASL 501) was added to achieve good accuracy (Figure 36). On average, it was reported that the limbus tracker was 37 ± 13ms faster in detecting the onset of saccades than the video-based eye tracker. The system was successfully applied in a saccade-contingent updating study for virtual reality researches.

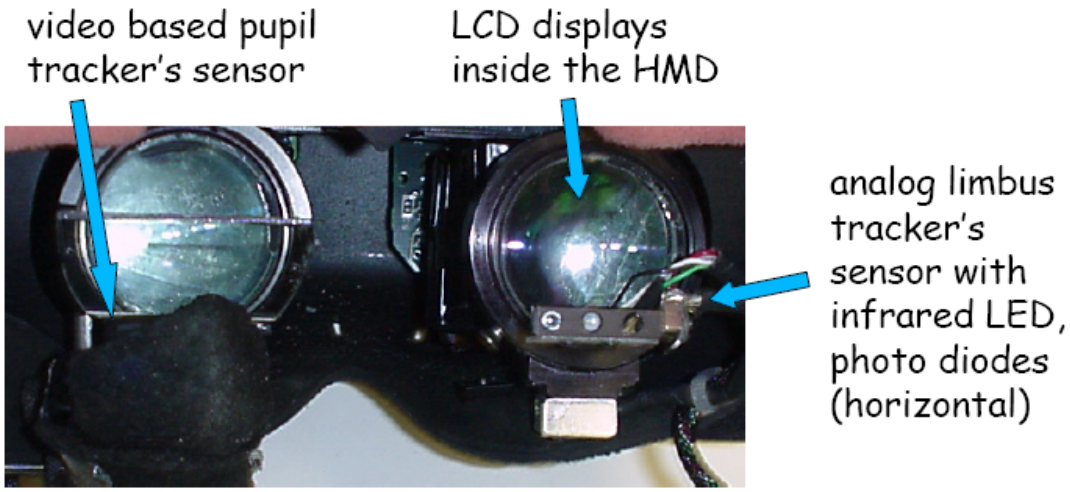

Figure 36. A hybrid system combining the limbus tracker and video-based tracker. Image from Jochen Triesch, 2002.

Other miscellaneous eye tracking techniques include: (1) observing the subject's eye directly. Although simple, an eye movement of one degree can be detected by an experienced experimenter (Yarbus, 1967, page 18); (2) scanning the fundus and then using the sequential fundus-retinal images to determine eye movements; (3) Interestingly, Ewald Hering in 1879 reported that an eye movement can be heard by listening to the muscle action through a stethoscope. During a fixation, dull clicks with

an interval occurrence are attributed to microsaccades, while steady background noises are associated with drifts and tremors (Carpenter, 1988, page 426).

For a further reading about eye tracking techniques, early surveys by Carpenter (1988) and Young & Sheena (1975), and recent reviews by Collewijn (1999), Duchowski (2002), Glenstrup & Engell-Nielsen (1995), Green (1992), and Richardson & Spivey (2004) provide much valuable information. An Eye Movement Equipment Database (EMED) maintained by Applied Vision Research Center also gives a good source for comparing the characteristics of different eye tracking systems.

5.7 Applications in Visual Perception Studies

Eye tracking techniques have been applied in a wide range of research, such as human computer interaction (HCI), clinical diagnosis and visual perception. The dynamics and characteristics of eye movements in a controlled laboratory have been studied extensively. One of the pioneering works was done by a Russian scientist, Alfred Yarbus, who recorded eye movements by attaching a "cap" containing a small mirror to the eyeball by suction (see Section 3.7.4). In the study, the subject was asked to examine a painting, "The Unexpected Visitor," given different instructions, including viewing freely, estimating the economic level of the family, telling the age of the people, guessing what the family was doing before the arrival of the visitor, remembering the people's clothes, recalling the positions of the people and objects in the room, and estimating how long the visitor has been away from the family. The different gaze patterns of the subject when given the different tasks demonstrate that viewing behavior is dependent on the ongoing tasks (Yarbus, 1967); human vision is closely associated with a cognitive process,

meaning an active system. This seminal work has been replicated by Lipps et al (2004) using a modern eye tracker, which is much less uncomfortable and requires less restriction of subjects' motion. Similar results were found even though the viewing time was quite different in the two experiments (self-paced about 20 seconds versus enforced 3 minutes in Yarbus).

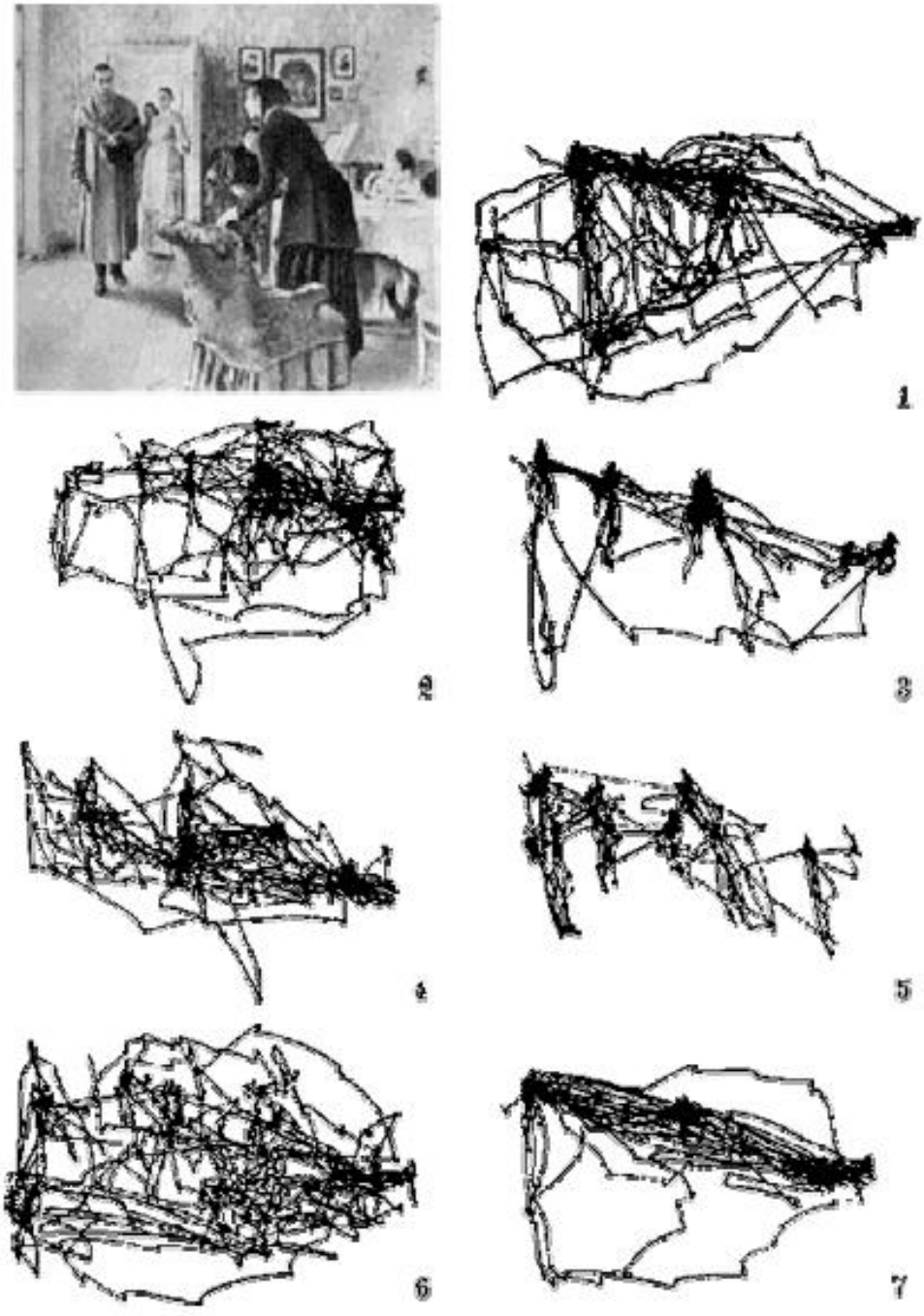

Figure 37. The painting of "The Unexpected Visitor" and the gaze patterns of the subject given different instructions. Image from Yarbus, 1967.

However, "outside the laboratory human vision is a tool, not a task" (Pelz & Canosa, 2001). Thanks to the advancement in portable eye trackers (Babcock & Pelz, 2004; Land & Lee, 1994; Pelz, 2004), today eye movement studies can be accomplished under more natural conditions, allowing unrestrained eye, head and body movements. Michael F. Land and his colleague (Land & Lee, 1994) worked on finding viewing strategies when a driver steers, using a head-mounted video system to simultaneously record both the road and the driver's eye. They found that drivers pay attention to the "tangent point" on the inside of curves; drivers search for this point 1-2 seconds before each bend and return to the point throughout the bend. Other studies for understanding how humans perform everyday extended tasks include making tea (Land, Mennie et al., 1999), hitting the ball in cricket (Land & McLeod, 2000) and hand-washing (Pelz & Canosa, 2001). The common results of these studies are that most fixations of the subject are tightly linked to task-relevant objects, occasionally being made to future objects in sequence. Such a fixation pattern, looking at the right place at the right time or "just in time," reveals how high-level perception and cognition are involved with our daily activities. These extended tasks are influenced very little by low-level cues in the scene such as the conspicuity or saliency of objects (Pelz & Canosa, 2001).

For further reading of eye movement studies in active vision, Hayhoe's papers (2004a and 2004b) provide reviews of current research directions.

Chapter 6 Eye Tracking Optimizations: Key Issues and Possible Solutions

Since the seminal work by Merchant and Morrissette (1974), video-based eye tracking techniques have been actively studied for more than thirty years. Although significant improvements in hardware and software have been made, there is still a lot of challenging work remaining. The first section of this chapter has an outline flavor, addressing some of the key issues in current video-based eye tracking systems. Extended discussions about those problems are provided in the rest of the chapter. The theory and description of a model-based approach, attacking the limited measurement range and artifact removal in the video-based eye tracking techniques are detailed in Chapter 7 and Chapter 8.

6.1 Key Issues in Video-based Eye Tracking

In this section, key issues in current video-based eye tracking systems are summarized and corresponding possible solutions are discussed.

The first issue is that of eye camera movement with respect to the head (translational eye movement). This is a significant problem in head-mounted video-based eye tracking system. Section 7.1 gives details about the definition of translational and rotational eye movements. In general, the techniques attacking the problem can be divided into two categories (Moore, 1999): (1) a better mechanical mounting of eye camera to the head, and (2) camera movement compensation by advanced image processing techniques. The techniques in the first category usually require some attachment to tightly fit the headgear to the head. Although less than 0.1mm slippage can be achieved, the method may not be suitable for application duration longer than 30min. Otherwise, discomfort may rise in the subject (Moore, 1999). Another promising approach is to have the subject put on a facemask made by thermoplastic materials (Figure 38). Low cost, minimal pressure points and relatively good performance make this method favorable. However, in order to have an optimal performance, a custom facemask is required for each subject.

Figure 38. Eye Tracking head unit with mounted facemask. Image from CHRONOS VISION GmbH, Germany. www.chronos-vision.de/eyetracking.

Other approaches, based on tracking additional features in the eye/face or

compensating for the camera slippage by advanced image processing techniques, have been reported:

(1) Combining tracking the pupil and CR. This is the technique on which most video-based eye trackers rely. As described in Section 5.5, by integrating the pupil tracking with the CR tracking, the confusion of translational eye movements can be reduced. This is because the positional difference vector (P-CR) between the pupil center and CR center changes primarily with rotational eye movements. However, when applying this technique, the CR easily rolls off the cornea in large eye movements, resulting in track losses. Furthermore, the use of the CR adds noise to final signals due to the CR having a much smaller size than the pupil. The small pixel counts of the CR in an eye image make its detection a relative poor process (Kolakowski & Pelz, 2006).

(2) Like the CR, other features, such as marks pasted on the skin, the upper eyelid or the canthus (corner) of the eye can also been tracked with the pupil for compensating the camera slippage.

(3) Based on the facts that the camera movements are much slower and less frequent than eye movements, the camera movements can be extracted in video-based eye tracking outputs (Kolakowski & Pelz, 2006; Li, Kolakowski et al., 2008). In this technique, the CR vector is only used to determine camera movements, which are smoothed and then subtracted from the pupil vector. Therefore, the final eye-in-head vector – equal to the pupil position vector minus the camera position vector – will represent rotational eye movements only with a noise level comparable to that in the less-noisy pupil position vector. The optical relationship about the CR and pupil offsets during translational and rotational eye movements, derived in Chapter 7, was used to compensate for the camera movements as shown in Kolakowski and Pelz (2006).

The cornea is a spherical surface only within its central 25 deg while it is modeled as a perfect sphere in most commercial and laboratory eye trackers. When the CR falls to the limbus (cornea-sclera interface), where the curvature flattens out, a large, uneven reflection may be observed (Figure 39). Furthermore, the variation of the CR between and within individuals, and the change of its shape when falling on the different parts of the eye (due to the different thickness and smoothness of the tear film) also cause problems in detection. This issue can be worse if the CR rolls off the cornea, over the white sclera, where it has a low contrast with the surrounding. In this situation, the tracking of the CR will become extremely challenging, if not impossible. In order to resolve the problem, either the cornea needs to be modeled by a more complicated shape (as described in Section 3.1) or measurement of eye movements have to be restricted to a limited range. In the presented manuscript, we propose another approach that relies on multiple illumination sources to produce multiple CRs, so as to keep at least one CR trackable, falling on the pupil or iris where it has a high contrast with the surroundings. The description of this new technique is covered in detail in Chapter 8.

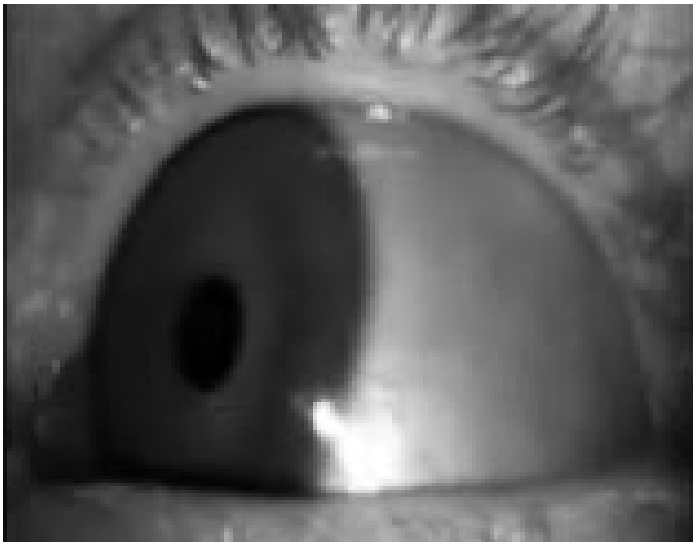

Figure 39. Observed irregular reflection when the CR falls on the cornea-sclera interface (limbus).

The real CR can be confused with spurious reflections in the eye image. These unwanted reflections can be from different sources, such as extra ambient illumination, glaring reflection on the glasses (or the frame), or reflection by other surfaces of the eye. If these artifacts appear in the image, additional criteria (like geometrical limitations) need to be used to sort out the real CR. However, in many cases the new criteria may not solve the problem completely; more than one specular reflection may still exist after the additional processing steps. If this happens, some eye tracking systems (e.g., ASL 6000 series) take the bright spot closest to the pupil as the desired CR (ASL, 2005). While this simplification makes the system run smoothly, it leads to errors and therefore degrades the accuracy of the system. Image classification techniques, by training the system with a large group of eye images containing artifacts and desired CR, may be an alternative way to solve the problem (Alex, Myron et al., 2001; Zhu & Ji, 2005). In order to differentiate the desired CR from spurious CRs efficiently and accurately, a model-based prediction

technique is proposed in this thesis, which takes advantage of the well-correlated relationship between the offset of the pupil and that of the CR (see Chapter 7 and Section 8.6 for details).

Infrared illumination is commonly used in video-based eye trackers. Though it is invisible to the eye, the detection of eye features may be influenced by other infrared light in the environment, especially outdoors under strong sunshine. Figure 40 shows an image taken outdoors on a sunny day during the summer in Rochester, NY. As shown in the figure, the eye features (the pupil and CR) are clearly visible - have high contrasts with their surroundings. The radiation of the IRED appears overwhelming in the total illumination because the on-board illumination source is mounted close to the eye. Also, because of the strong illumination in the outside environment, the size of the pupil is significantly smaller than it appears in an indoor scene. Furthermore, the illumination level of the eye image may vary greatly due to the occurrence of uncontrolled conditions, such as walking into the shadow of a tree. Therefore, a good adaptive threshold algorithm is needed to isolate the pupil and CR according to different illumination conditions. Another approach worth trying when working in complex illumination environments is appearance models (Tan, Kriegman et al., 2002), in contrast with the above feature-based methods. The training approach mentioned previously may also be a feasible solution.

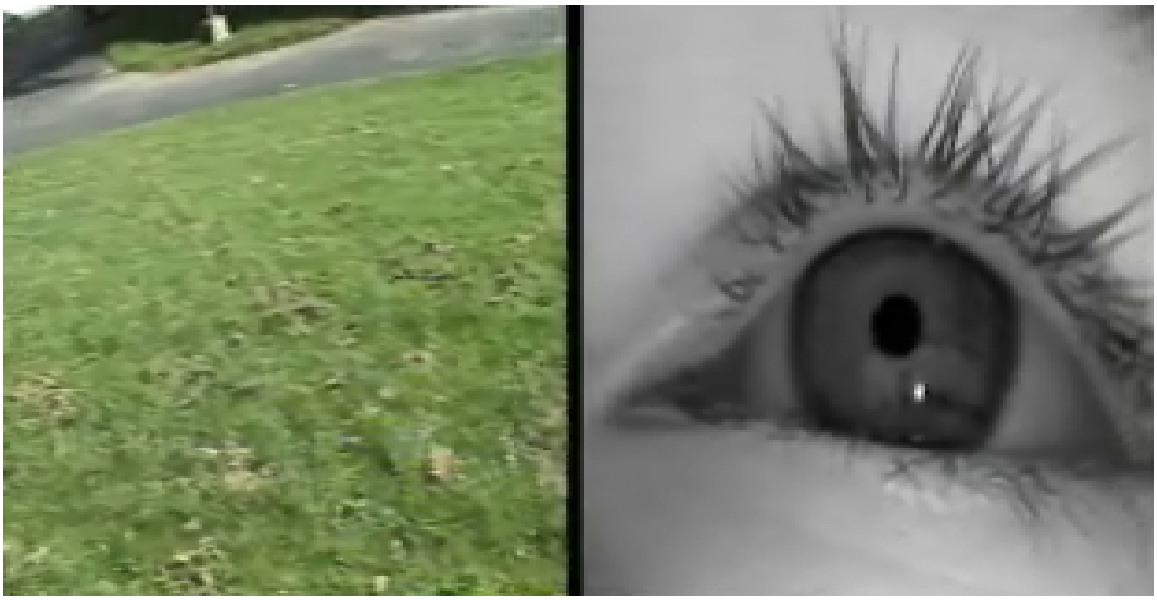

Figure 40. Scene (left) and eye (right) images taken outdoors on a sunny day.

Parallax errors can be introduced if the calibration plane and fixation surfaces are different. Calibration is used to determine the parameters for mapping the eye positions in head coordinates to world coordinates (i.e., scene or display coordinates). If the eye tracker is calibrated in a plane that is different from where the subject is fixating, a parallax error will be introduced. As shown in Figure 41, for the case that the distance between the eye and scene cameras is 1.5 inches, when the calibrated and fixated planes have equal distances to the eye, the parallax error is equal to zero (darkest gray region on the diagonal). Therefore, whenever possible in an eye tracking study, the visual stimuli should be presented on the plane having the same distance away from where the calibration is conducted. If the stimuli would be displayed in variable distances from the eye with known minimum d_{min} and maximum d_{max} working distances, the optimal calibration distance is found to be $d_{min}+(d_{max}-d_{min})/3$ (Li, 2006).

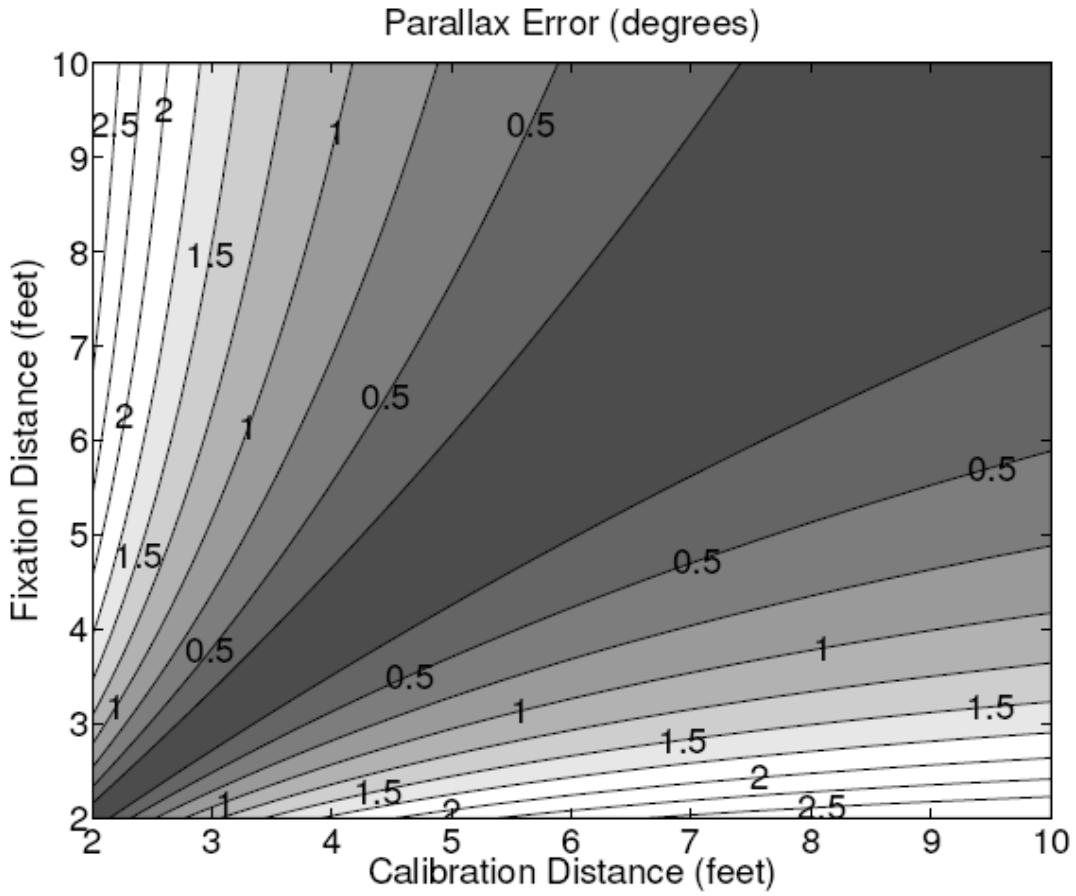

Figure 41. Parallax error for the case when the distance between the eye camera and scene camera is 1.5 inches. From (Li, 2006).

The low sampling rate of the eye cameras is a bottleneck for recording fast-phase eye movements (e.g., saccades). An intuitive way to overcome this bottleneck is to apply high-speed cameras in capturing eye images (see Section 6.2). Existing high-speed eye trackers with a sampling rate of 1000Hz or above are usually bulky (e.g., EyeLink 1000 by SR Research Inc.) and use custom video cameras (technical details of which are considered as proprietary). An alternative strategy is to model stereotyped eye

movements. As described in Section 4.2, “The Main Sequence” of saccadic movements can be represented by a simple equation (Equation 4-1). Taking into account the dynamics of the eye movements, their trajectories can be predicted and reconstructed with a few sampling points, which compensates for the limited sampling rate of the eye camera.

The discrepancy between the eyeball axis and head axis: the eye position in space cannot be obtained by simply summing the eye position in head and head orientation. A more advanced technique is necessary to accurately derive eye orientations in space.

The deviation between the visual axis and optical axis of the eye: this angular discrepancy is about 5° on average (see Section 3.5) and needs to be accounted for when calculating the eye position (though the calibration step may compensate partially for this offset).

A result that has both negative and positive implications is the voluminous stream of eye tracking data. In many eye tracking studies, an analysis needs to be done by frame-by-frame examinations (data are usually recorded on video tapes), though some software and algorithms are developed to aid the analysis (Gitelman, 2002; Juhola, Jantti et al., 1985; Munn, 2009; R.Tole. & R.Young, 1980; Salvucci & Goldberg, 2000). Relying on the understanding of eye movements, and advanced techniques in signal and image processing, smart software can be built to alleviate the burden of analyzing the voluminous eye tracking data.

Except for the camera motion with respect to the head (an apparent translational eye movement seen in the eye image), the translation of the eyeball within the socket can also be an influencing factor. Fortunately, it is believed that this effect is not significant and, as such, can be neglected (Borah, 1989).

6.2 Cameras for High Speed Imaging

As stated earlier in this chapter, current video-based eye tracking systems are commonly limited in the sampling rate of eye cameras when recording fast-phase eye movements. Using a low speed camera to capture a fast moving target (e.g., fast-phase eye movements or visual stimuli) will cause blurred images, or even worse, multiple images (Figure 42). In some studies, such as on saccades or fast-phase nystagmus (up to 600°/s), 250Hz sampling rate is the minimum requirement (Karn, 2000). Today, many eye tracker manufacturers produce systems capable of a sampling rate of 1000Hz or more, such as iViewX Hi-Speed (1250Hz) by SMI (SMI), EyeLink 1000 by SR Research (SR_Research) and Chronos 1000Hz by Chronos Vision (Chronos-Vision). High speed eye trackers (< 1000Hz) are also offered by some other traditional eye tracking companies, such as Applied Science Laboratories (ASL) (up to 360Hz), ISCAN (up to 240Hz), SR Research (EyeLink II, up to 250Hz when recording both the pupil and CR simultaneously). These products give customers flexibilities when dealing with high-speed motions, whereas most of them, if not all, use custom high-speed cameras and keep technical details proprietary.

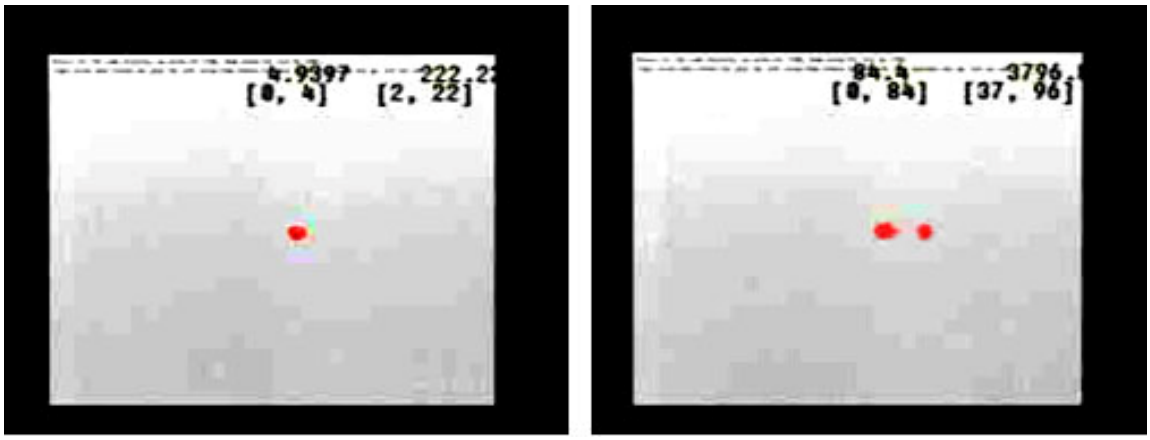

Figure 42. A target moving at 222 pixels/s (left), and 3796 pixels/s (right) is captured by a standard 60Hz camera. The multiple targets seen in the right image is caused by using a slow 60Hz NTSC camera to capture the fast moving target.

Another advantage of using a high speed camera is that better noise reduction can be achieved by averaging multiple continuous fields. For example, averaging six fields in a 360Hz high speed camera is equivalent to the same sampling rate of a standard NTSC camera (60Hz), but the process can dramatically reduce noise in the output.

Improvements in semiconductor and computer technologies make it possible for high-speed eye trackers to provide high quality images with a modest cost. Two kinds of solid state sensors are commonly used in high-speed cameras, CMOS and CCD. CCD used to be dominant in the imaging sensor market because of its high sensitivity and high quality image output. However, a high-speed camera design using CCDs is limited in several aspects, such as the requirement of a special fabrication process (high cost), blooming (overexposure) and smear (unwanted stripes), and serial readout (relatively slow output). The advent of CMOS techniques makes these problems relatively easy to overcome. High integrated circuitry and low power consumption are also superiorities of CMOS sensors over CCD sensors in mobile applications.

Many companies market CMOS sensors capable of 1000fps at full-frame rate (Redlike, Vision Research, Weinberger, etc.), even 3000fps (Photron), while most of them use their own proprietary CMOS sensors and package them in camera sets. Off-the-shelf high-speed CMOS sensors are also available from several companies, e.g., Micron (now Aptina) and Cypress. Table 6 lists some primary parameters of two off-the-shelf high-speed CMOS sensors.

Table 6. High speed CMOS sensors from Micron and Cypress

	MT9M413	LUPA-1300
Manufacturer	Micron (Aptina)	Cypress
Resolution [pixel×pixel]	1,280H x 1,024V	1280H x1024V
Pixel size	12.0μm x 12.0μm	14.0μm x 14.0μm
Full frame rate (fps)	500	450
Sensor imaging area	15.36mm x 12.29mm	17.9 mm x 14.3 mm
Subwindow capability	vertical change	programmable
A-D (bit)	10	16
Dynamic Range (dB)	59	62
Color or monochrome	either	either
Spectral sensitivity range	400 – 1000 nm	400 – 1000 nm

A high-speed eye tracker provided by ASL works at up to 360Hz (6 times faster than standard NTSC cameras). In a high-speed mode (240 or 360Hz), the camera skips scanning the top and bottom of the field which results in decreasing the field of view (a

decrease to 1/2 of its original array if working at 240Hz, or 1/3 at 360Hz). As a result, the vertical measurement range is reduced significantly, which may not be acceptable in some eye tracking studies. The pupil or CR likely moves out of the field of view of the camera as the subject looks toward the vertical extremes. This tradeoff also makes the positioning of the eye camera in the vertical direction very critical. The horizontal limit is kept the same in the camera design. Furthermore, a high frame rate requires a short exposure time, which demands a strong illumination of the eye, subdued ambient lighting and setting adjustment in the image processing. The camera gain may also need to be tuned to optimize the output.

Another high-speed eye tracking approach is reported by Talukder, Morookian et al. (2005), who proposed an eye tracking architecture that was capable of operating at speeds as high as 6~12kHz. The system relies on a CCD sensor (model TC237 by Texas Instruments) that has a standard 60fps at the full frame rate with sub-windowing capability. When the camera searches for the pupil and CR, it operates at the full-frame mode. Once the features have been detected, the camera only captures the sub-region around the pupil in the previous frame, resulting in a rapid updating rate. Although the sub-window is refined on a frame-by-frame basis, the system may still be subject to the fast motions of the eye, as the eye features, once lost, may take a while, or involve a manual operation, to be relocated.

A group of researchers in Caltech is building a real-time, event-driven (RARE) camera for NASA's Jet Propulsion Laboratory. This CCD sensor is designed to have a frame rate of 6 kHz (outputting a 8×8 pixel sub-window of an image) and to be adaptable to eye tracking applications (NASA, 2002, 2004).

6.3 Proposed Illumination Configuration

One of the serious problems in video-based eye tracking is the track loss when the CR rolls over the sclera (see Section 6.1), where the CR has a degraded contrast with its surrounding. What we term "structured illumination" utilizes an array of infrared sources or multiple beams out of one single source to generate multiple CRs. This ensures, even during an eccentric gaze, at least one trackable CR in the eye image; the CR falls on the pupil or iris, where it has a high contrast with the background. The use of multiple CRs avoids the track loss during large eye movements and hence extends the measurement range of the eye tracker. Moreover, the noise level of a computed CR vector may be reduced dramatically by averaging or median-filtering the positions of all trackable CRs. A conceptual illustration of the structured illumination is shown in Figure 43.

Figure 43. A conceptual illustration of the structured illumination (left) and the highlighted CRs and pupil (right).

When using multiple illuminators, differentiating CRs to compute the P-CR vectors turns out to be crucial and challenging. Two approaches have been proposed, and their advantages and disadvantages are discussed in the following two sub-sections. The last sub-section introduces a novel approach that does not require the differentiation process,

but rather relies on a model-based CR prediction technique.

6.3.1 Multiple CRs with Temporally Sequential Illumination

From the perspective of hardware implementation, temporally sequential illumination is relatively easy to realize. By synchronizing the onset of each illuminator with image capture, all CRs can be detected and determined individually in an illumination period (a cycle of onsets of all illuminators, e.g., one period is 0.2s if using six illuminators with a standard NTSC camera of 30fps). Followed by an averaging or voting (e.g., median-filtering) stratagem, an "optimal" CR can be selected at each illumination period. This technique, however, has a limited temporal resolution because of the low sampling rate of video cameras (commonly 60Hz). Another issue in this approach is that the system may not be fast enough to process (vote or average) the CRs during a rapid eye movement.

If a continuous device is used instead of a discrete sensor, the bottleneck of the sampling rate can be solved due to the large bandwidth of continuous devices. Such an idea has been implemented in a HiBall Tracking System (Greg, Gary et al., 1999; Welch, Bishop et al., 2001). The system (Figure 44) consists of a group of head-mounted sensors called HiBall, a fixed infrared LED array called Ceiling, and a circuitry board for communication and synchronization called CIB. The LEDs are flashed sequentially (one at a time) and their luminous flux is collected by lateral effect photodiodes (also called position-sensing detectors, devices that output the centroid of the incoming luminous flux) in HiBall. The HiBall also converts optical signals to the orientation information of the user's poses. The system, generating over 2000 pose estimates per second, has less

than one millisecond of latency, better than 0.5 millimeters resolution, and 0.03 degrees of absolute error (Welch, Bishop, et al., 2001).

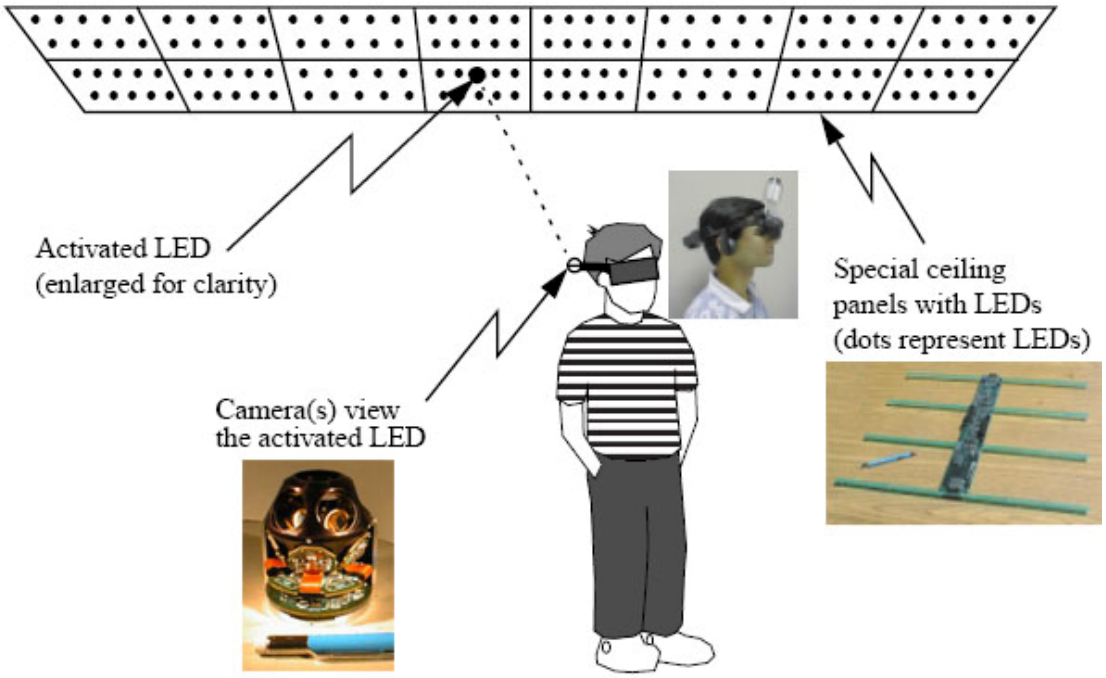

Figure 44. The HiBall Tracking System. Adapted from www.cs.unc.edu/~tracker/.

6.3.2 Multiple CRs with Multispectral Imaging

An alternative method to distinguish different CRs is to use multiple illuminators with different peak wavelengths (like IREDs). The multispectral imaging technique in the remote sensing community can be borrowed to differentiate the CRs. A diagram shown in Figure 45 illustrates how to detect four infrared rays with different wavelengths. Thin-film filters can be added in front of each sensor to selectively admit certain bands of infrared rays.

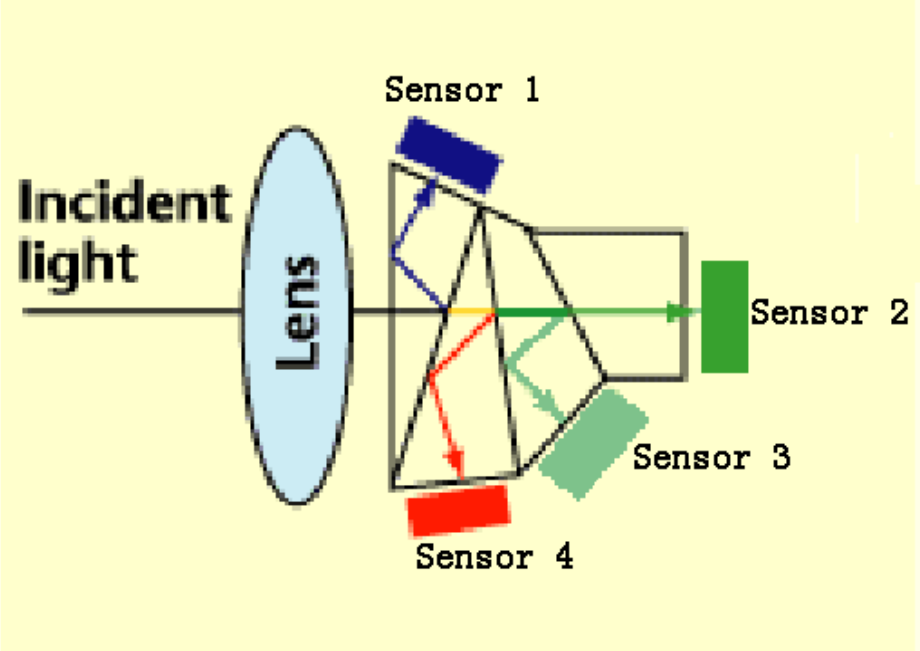

Figure 45. A band separation prism with four optical sensors.

This method keeps the same temporal and spatial resolutions in all four images captured by the sensors, but if a number of images with different wavelengths would be captured, the system might become very bulky.

6.3.3 Multiple CRs with CR Prediction Technique

All the methods described above aim to distinguish the images of multiple illumination sources (CRs) in each eye image. In this section, an approach that does not rely on the discrimination of the CRs is discussed. As shown in Chapter 7, the pupil and CR both move in the direction of the eye movements, but with different offsets ($Pupil_{offset} : CR_{offest} \approx 2:1$). Based on this well-correlated relationship, the CR positions can be predicted if the

pupil positions in the previous and current frames and the CR positions in the previous frame are known (Figure 46). Furthermore, genuine CRs can be separated from spurious CRs by searching around the predicted CR regions if there are no significant camera movements. The detailed derivation of the relationship between the offset of the pupil and that of the CR is given in Chapter 7. The implementation of this prediction technique in video-based eye tracking is presented in Chapter 8.

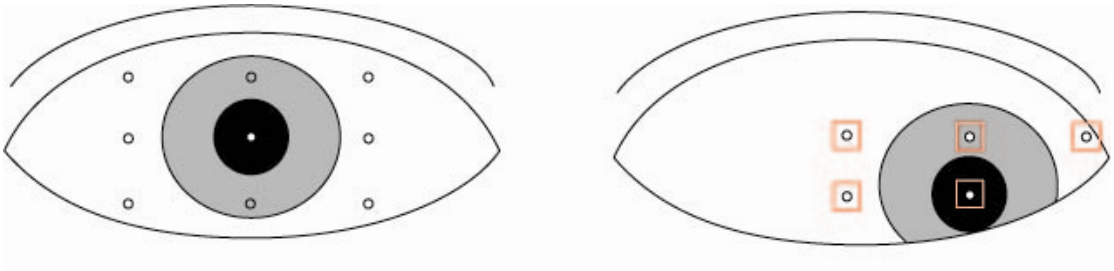

Figure 46. CR-prediction technique. Left: original eye position. Right: eye position after a rotational eye movement. Red squares represent the predicted CR positions.

On the other hand, this technique can also be used to detect (and quantify) camera movements because the actual CR positions and the predicted ones will not match after the eye translates (Figure 47) if the camera has moved.

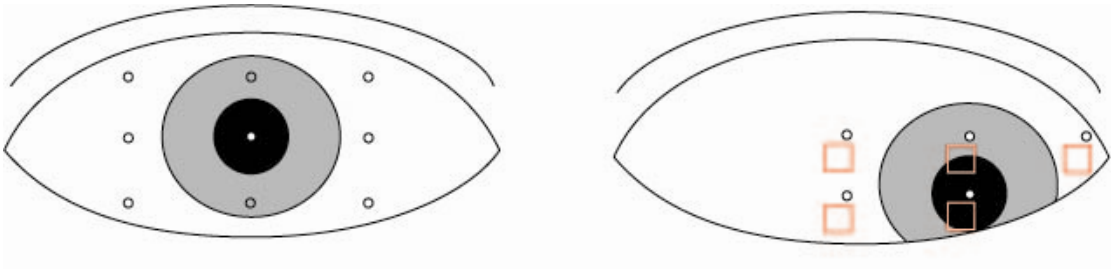

Figure 47. After a camera movement (translational eye movement), the predicted positions of the CRs and the actual positions do not match. Left: original eye position. Right: eye position after a translational movement.

6.4 Optimizing Illumination Collection

Most video cameras sample the eye position at 60Hz, which is not fast enough to capture the full range of eye movements (e.g., saccades' velocity up to 600°/s). High-frame-rate cameras can overcome this problem, but they require a short exposure time in a fast speed mode, which may degrade the quality of captured images. For cases in which a telephoto (high magnification) optical lens is used (e.g., in some remote eye trackers), the situation becomes even worse because of the further reduction of light striking the imaging sensor plane (Beutter, Borthwick et al., 1998). Increasing the intensity of the illuminator can be a solution. From the perspective of safety and comfort to the subject, however, the illumination level should be kept below 10mW/cm² (Clarkson, 1989; Sliney & Wolbarst, 1981). Lower values are better. These two contradicting requirements can be balanced partially, if not completely, by carefully choosing optical components for image capture.

Taking the RIT Lightweight Eye Tracker (Figure 48) as an example, if we would want to apply a high-speed camera, several aspects should be taken into account when building an optical module. The first and most important of these is to choose a high-frame-rate imaging sensor that is sensitive to infrared illumination. In order to improve the detection of the desired eye features (pupil and CR), a composite lens, mounted in front of the sensor, should be coated to be transparent in the infrared wavelength and block the visible spectra. It is also necessary to block wavelengths beyond 1100nm to avoid potential hazards. The second step is to choose an appropriate infrared illumination source. An IRED (or IREDs if using multiple illuminators) should

emit radiation to which the eye camera is most sensitive. The illumination angle of the IRED must also be chosen appropriately in order to get an evenly illuminated eye image. For an illumination configuration similar to the RIT Eye Tracker, an IRED with ± 20~30° angle should be wide enough since the light source is mounted close to the eye. For optical configurations like those in ASL and ISCAN eye trackers (see Figure 30 and Figure 33), the visor (located in front of the eye) needs be coated to reflect the infrared illumination and transmit the visible spectrum.

Figure 48. Headgear for the RIT Lightweight Eye Tracker

Chopped light (light modulated at a certain frequency) can be a good option to reduce confusion with the ambient illumination and improve the detection of the IR

illuminator shining on the eye. When using an IR source in an eye tracker, the safety limit of 10mW/cm² should always be kept in mind (Clarkson, 1989; Sliney & Wolbarst, 1981).

6.5 Filtering Strategies in Minimizing Noise

Even with an optimal optical configuration and illumination collection, a captured image will still contain noise and involve other issues. In order to gain the maximum information from an eye image or the eye tracking data, filtering strategies are often necessary.

a) Noisy P-CR vector

The P-CR vector, which is used to determine the line of sight of the subject in video-based eye tracking, is the displacement vector between the pupil center and CR center. Due to the inclusion of CR, the P-CR vector has a noise level no better than the CR vector (Kolakowski & Pelz, 2006). The latter is relatively noisy due to the smaller size of the CR in the eye image and its smaller displacement during eye movements in comparison with the pupil motion. If only the pupil vector would be used to determine the line of sight, instead of the P-CR, then the signal-to-noise ratio in the eye position data would be greatly improved. A promising approach reported by Kolakowski et al. (2006) uses the CR vector only to extract the camera motions (translational eye movements). The calculated eye-in-head position is equal to the pupil-position vector minus the camera-position vector. This approach achieves a noise level in the eye gaze outputs comparable to that in the less-noisy pupil positional data while it is still capable

of detecting small, rapid eye movements (Kolakowski & Pelz, 2006).

b) Blinking

In most video-based eye tracking systems, blinking is unavoidably included in the recorded outputs. After plotting the point of regard (POR) data, blinks are always observed to be extreme or special values in the output data (such as in the ISCAN eye trackers). By a simple thresholding operation, a majority of blinks can be removed as shown in Figure 49.

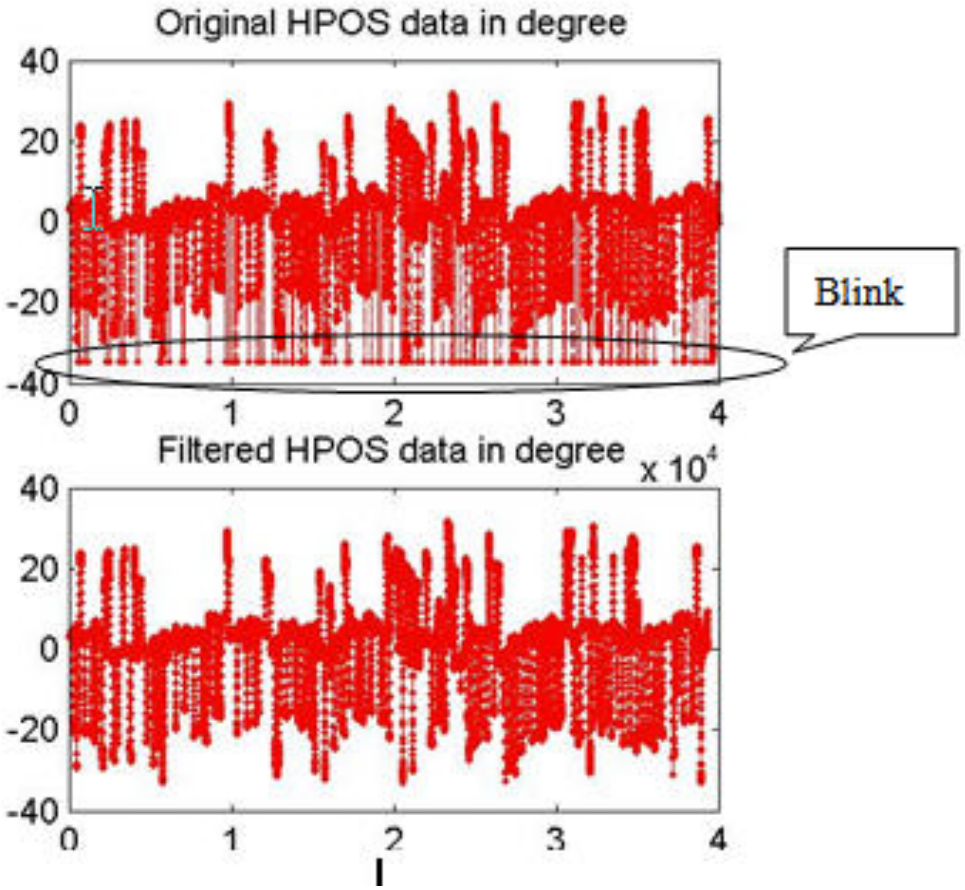

Figure 49. Blinks removal by thresholding. Original data are outputs from an ISCAN eye tracker. HPOS is the horizontal position of the point of regard (POR).

After detecting the blink, the observations in those positions can be replaced by the interpolated values based on the data before and after the blink. The interpolation operation is better than a zero-padding process because the blink may happen during an eye movement; the zero-padding step may split a single fixation into multiple segments.

Another approach, based on the circularity evaluation of the detected "pupil" region, is introduced in Section 8.5.

c) Successive odd and even field jitter

For interlaced videos, jitter is a common issue caused by the mismatch of odd and even fields in one image frame. A simple solution is to de-interlace the problematic frame into two fields, an odd field and an even field. Then a correlation operation is applied to calculate the offset between the two fields. The deviation between odd and even fields can be eliminated by compensating for the offset.

Another approach is by image averaging as having been applied in some commercial eye trackers (e.g., ASL 6000 and ISCAN eye trackers). By averaging multiple continuous frames, the jitter (and other noises) can be reduced. However, this method sacrifices temporal and spatial resolution.

d) Motion blur

A blur image is frequently seen from outputs of video-based eye trackers due to headband slippages (for a head-mounted type) or fast-phase eye movements. Motion de-blurring techniques, such as the Wiener Filter or the Lucy-Richardson Filter, can be adopted to alleviate the problem.

The motion blur is modeled as a convolution process:

$$b(x,y) = psf(s,t) * o(x,y) + n(x,y)$$

where $b(x,y)$ is the blurred image, $o(x,y)$ is the original image, $n(x,y)$ is the noise, and $psf(s,t)$ is the point-spread function of the blurring process, which can be spatial and/or temporal-dependent.

In the Fourier domain, the model is described as:

$$B(k_x, k_y) = MTF(k_{x,t}, k_{y,t})O(k_x, k_y) + N(k_x, k_y)$$

In the noise-free case, when the blurring function is known, the de-blur process can

be implemented by multiplying the inverse of the MTF function with the blurred image in the frequency domain.

Since it is hard to know the MTF function of the motion blur in video-based eye tracking, blind de-convolution techniques may be useful in reducing motion blur in the image (Kundur & Hatzinakos, 1996).

e) Spatial noise

Isolated noise, such as salt and pepper noise, can be caused by defects in an imaging sensor (e.g., hot and dark pixels). In most cases, a smoothing operation (e.g., median filtering) may effectively reduce that noise. Morphological operations, dilation followed by erosion-filling holes and erosion followed by dilation-erasing islands, are also good alternatives.

Chapter 7 Eye Tracking Optimizations: Modeling of Optical Relationships between Pupil and CR Offsets

Since video-based techniques track the pupil and CR to determine eye orientation, it is beneficial to understand the relationship between these two features during different types of eye movements; these relationships will be manifested in two values that we term the translational and rotational gains[♣]. The utilization and benefits of these values will be made clear in Section 8.6. In this chapter we first clarify translational and rotational eye movements, and then derive two gain values related to these two eye movements. The *translational gain* refers to the amount the CR moves when the pupil center moves one unit due to a relative translational movement between the eye and the eye camera. The *rotational gain* refers to the amount the CR moves when the pupil moves one unit due to a rotational eye-in-head movement. In order to enhance detection of the eye features (pupil and CR), near infrared (IR) beams are commonly used because they are invisible to humans and hence will not disturb or distract the observer. Collimated and near-field illuminations are two conventional approaches for illuminating

[♣] These two gain values are first defined as Eye Gain and Camera Gain in Kolakowski, S., & Pelz, J. (2006).

the eye. In the former approach, parallel beams are either emitted from a distant light source, which is approximately the situation for a remote desktop eyetracker, or modulated by optical components, which is usually seen in head-mounted eyetrackers (e.g., a popular commercial eyetracker by Applied Science Laboratory shown in Figure 30). In other illumination schemes, such as with the ISCAN eyetracker (another widely used commercial eye tracking system) and the RIT Eyetracker (a lightweight, wearable eye tracking system, Figure 35), the light source is mounted close to the eye (typically 50mm or less from the eye). We start by deriving rotational and translational gain values for the case of collimated illumination. Following this derivation, we present a more general derivation for near-source illumination which can be extended for an illumination source at any distance from the eye. A portion of the contents of this chapter were published in Li et al. (2008).

7.1 Rotational and Translational Eye Movements

With the increased flexibility of head-mounted eye trackers come artifacts in output of eye-in-head orientation that occur as the subject is freed to move through the environment. These artifacts arise as a result of translational movements of the eye-tracking camera with respect to the subject's eye, which produce displacements of the pupil and CR in the eye images that are mistakenly attributed to eye-in-head rotational movements. If the headgear were perfectly stable and moved exactly with the subject's head, and the eye camera and scene camera were securely mounted onto the headgear, the cameras would not move with respect to the subject's eye and any displacement of the pupil or CR within the eye images would be due solely to rotational

eye movements. Unfortunately, this idealistic scenario does not exist. The headgear may move a small amount (e.g., bob up and down as the subject walks or when the subject simply makes a facial expression), and small movements of the headgear have a large effect on the output eye-in-head orientation. A 0.1 mm movement of the eye camera with respect to the eye will introduce an artifact of about 1° in the output eye orientation (Carpenter, 1988, p. 415 & 416). As the eye camera moves with respect to the eye, the pupil and corneal reflection move within the eye image; if these pupil and CR movements are not discerned and compensated, they may be mistaken as apparent eye movements in the output data. For this reason, it is very important that camera movements be distinguished from eye movements using the available pupil and CR data.

For brevity, from this point on, we will refer to these two movements often "seen" by the eye tracker as *rotational movements* and *translational movements*, respectively, where rotational movements refer to rotational eye movements in the head and translational movements refer to translational eye movements with respect to the eye camera (see Figure 28).

The Gullstrand Number 2 Simplified Eye (see Section 3.6.2) is used in deriving all optical relationships in this chapter. Note this model may not provide the same parameters as does the physical eye, but it describes the optical performance of the eye with a fair amount of accuracy and is sufficient for the purpose of this research.

Considerable variations (~10%) exist (Charman, 1994) between individuals in surface curvatures, component separations and axis lengths; for this reason, the following optical derivations are representative of an average adult eye.

7.2 Collimated Illumination

7.2.1 Translational Gain

Without losing generality, the original eye position is assumed to be centered so that the CR is situated at the focal point of the cornea (see Section 3.7.3). As seen in Figure 50, for collimated illumination, the CR remains on the focal point of the cornea when the eye is translated from A to A′ (equivalent to a translational camera movement, with respect to the eye, in the opposite direction). The pupil center also has a translational displacement that is equal to the eye translational displacement. Therefore, when the eye translates with respect to the camera, the pupil center and CR move the same amount. In other words, the translational gain, g_{trans}, is unity as shown in the following equation:

$$g_{trans} = \frac{\Delta c}{\Delta p} = \frac{x}{x} = 1.0 \qquad (7\text{-}1)$$

where x, Δc and Δp are the eye translational displacement, the offset of the CR and the offset of the pupil center, respectively. Based on this result, video-based eye tracking techniques utilize the P-CR vector so that, in the case of collimated illumination, translational eye movements do not contaminate the measure of eye angular orientation. For large camera movements, the advantage of using the P-CR technique may be diminished because the cornea becomes aspherical beyond its central region, resulting in a bias introduced by the spherical model of the corneal surface. In order to get an accurate g_{trans} values for large eye movements, a more complicated model, such as those described in Section 3.1, is needed.

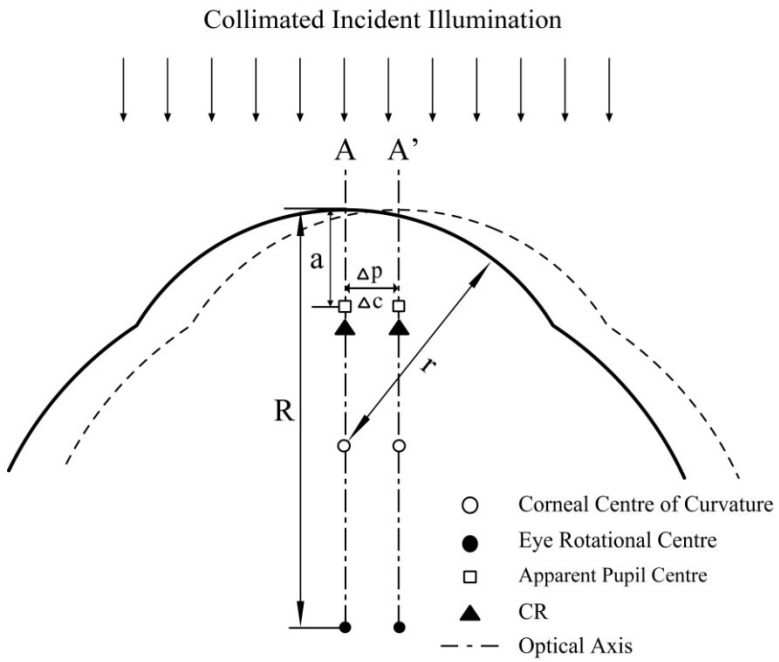

Figure 50. Eye before (solid) and after (dashed) a translational eye movement. The CR and pupil center move the same amount as the translational eye displacement if the incident rays are collimated; in this case, the translational gain is equal to 1.

7.2.2 Rotational Gain

When the eye rotates to a new position A′, the CR is situated at the intersection point of the focal plane of the cornea and the ray, in the direction of the illumination, going through the corneal center of curvature (see Section 3.7.3) as shown in Figure 51.

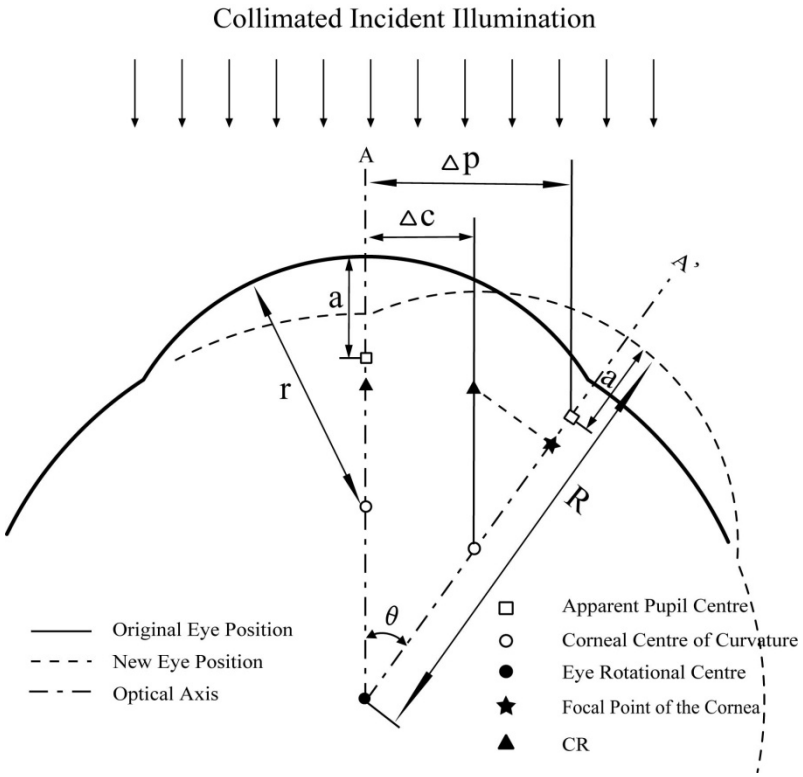

Figure 51. Eye before (solid) and after (dashed) a rotational eye movement. The rotational gain is equal to 0.55 if the incident illumination is collimated. Note that the CR after the eye rotates is located in the focal plane of the cornea in the new position.

The offset of the pupil center and that of the CR are represented in the following equations:

$$\begin{aligned} \Delta p &= (R-a)\sin\theta \\ \Delta c &= (R-r)\sin\theta \end{aligned} \qquad (7\text{-}2)$$

where θ is the rotational angle of the eye.The rotational gain, g_{rot}, is the ratio of these two offsets which is computed (using values in Table 3) as:

$$g_{rot} = \frac{\Delta c}{\Delta p} = \frac{(R-r)\sin\theta}{(R-a)\sin\theta} = \frac{(R-r)}{(R-a)} \approx 0.55 \qquad (7\text{-}3)$$

Based on this value, the CR moves about half the distance as does the pupil center during

an eye rotational movement because the pupil center is located at approximately twice the distance of the corneal center of curvature from the eye rotational center.

7.3 Near-Source Illumination

In the near-source illumination condition, the translational and rotational gain values derived in the previous section are no longer valid. In order to derive new values for near-source illumination, we make two assumptions:

- The CR is located in the focal plane of the cornea;
- The illumination source is a point source.

The first assumption is only true for a distant illumination source (see Section 3.7.3). However, for the typical usage with near-source illumination (i.e., source mounted about 30-40mm away from the eye) in the RIT Eye Tracker and given the fact that there are considerable variations between individuals' eyes (~10%) (Charman, 1994), this approximation is reasonable as we see the experimental validation in Kolakowski & Pelz (2006) and Li et al. (2008). The second assumption was already used when we calculated the apparent pupil center in Section 3.2 and the positions of the CR in Section 3.7.3. Without losing generality, the axis of the illuminator is assumed to be positioned directly in line with the optical axis of the initial eye position in our derivations.

7.3.1 Translational Gain

In a near-source illumination design, after a translational movement, the CR will not lie at the focal point of the cornea except when the line of sight goes through the illuminator

(original eye position in Figure 52).

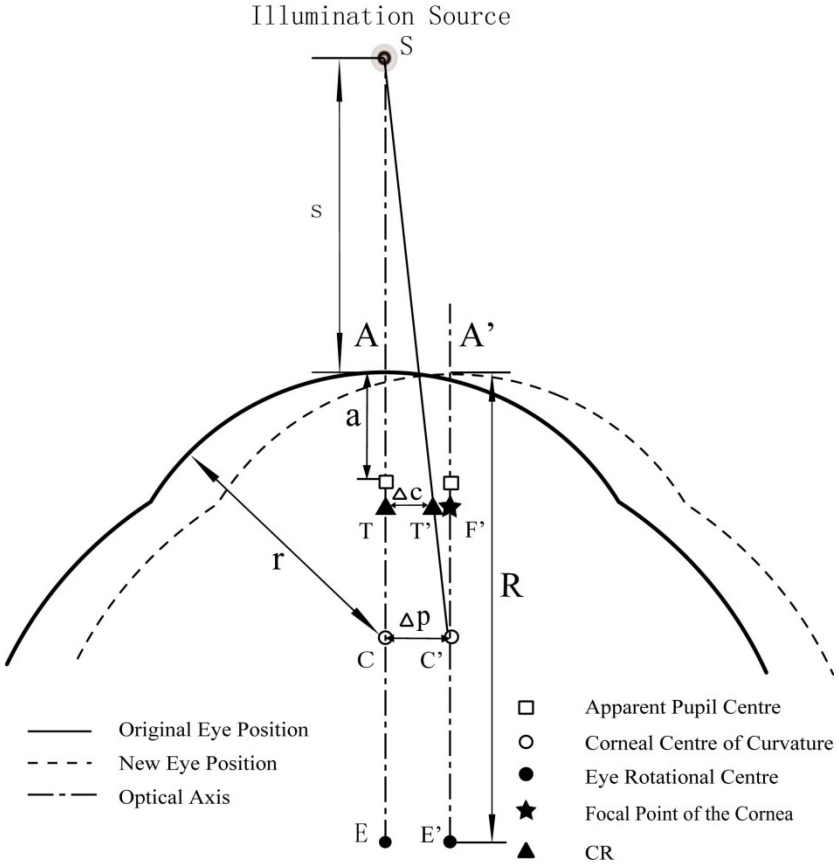

Figure 52. Eye before (solid) and after (dashed) a translational eye movement when the illumination source is close to the eye.

As before, the position of the CR after an eye translation is estimated as the intersection point of the ray passing through the corneal center of curvature with the focal plane of the cornea. Because the triangles STT′ and SCC′ are similar, as shown in Figure 52, the translational gain for any position of an illuminator is calculated as:

$$g_{trans} = \frac{\Delta c}{\Delta p} = \frac{r/2 + s}{r + s} = \frac{3.9\text{mm} + s}{7.8\text{mm} + s} \qquad (7\text{-}4)$$

For s equal to 30mm and 40mm (approximate limit values in implementations of

the RIT Eye Tracker), g_{trans} is equal to 0.90 and 0.92, respectively. These two values are significantly lower than unity, the translational gain in the case of collimated illumination.

Figure 53 shows the translational gain plotted as a function of the position of the illumination source using Equation 7-4. This plot indicates that using the P-CR vector cannot fully eliminate the contamination of translational eye movements unless the illumination source is far away from the eye (or the rays are collimated by additional optics). As the source is moved further from the eye, the translational gain approaches unity, which is the value under the condition of collimated illumination. Therefore, the translational gain in Section 7.2.1 can be considered as a special case of that derived in this section (Equation 7-4). This equation can be used to determine how far away a remote tracker must be in order to use the P-CR method based on the accuracy desired.

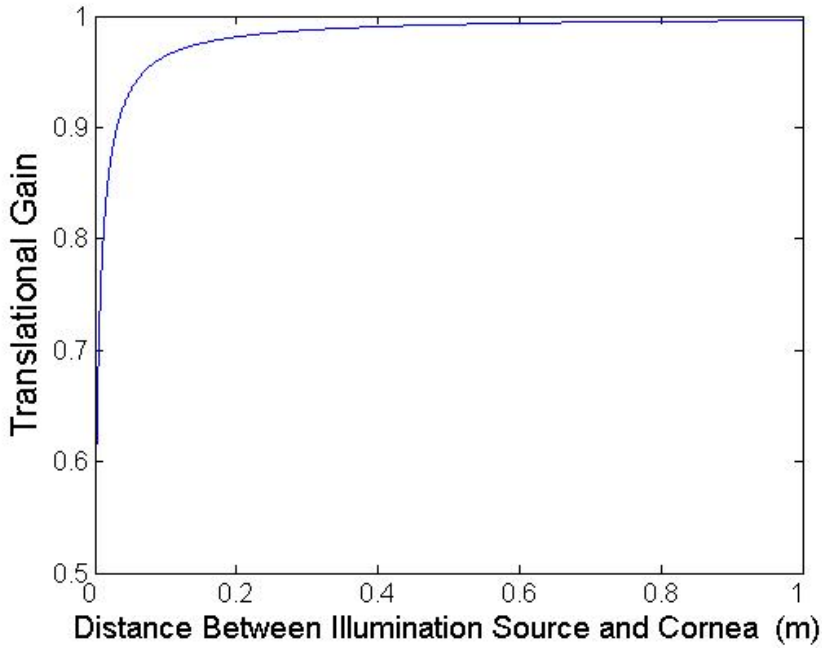

Figure 53. Translational gain versus position of the illumination source (with respect to the eye).

7.3.2 Rotational Gain

As in the previous section, the CR position after the eye moves is estimated as the intersection point (T′ in Figure 54) between the focal plane of the cornea and the ray passing through the corneal center of curvature.

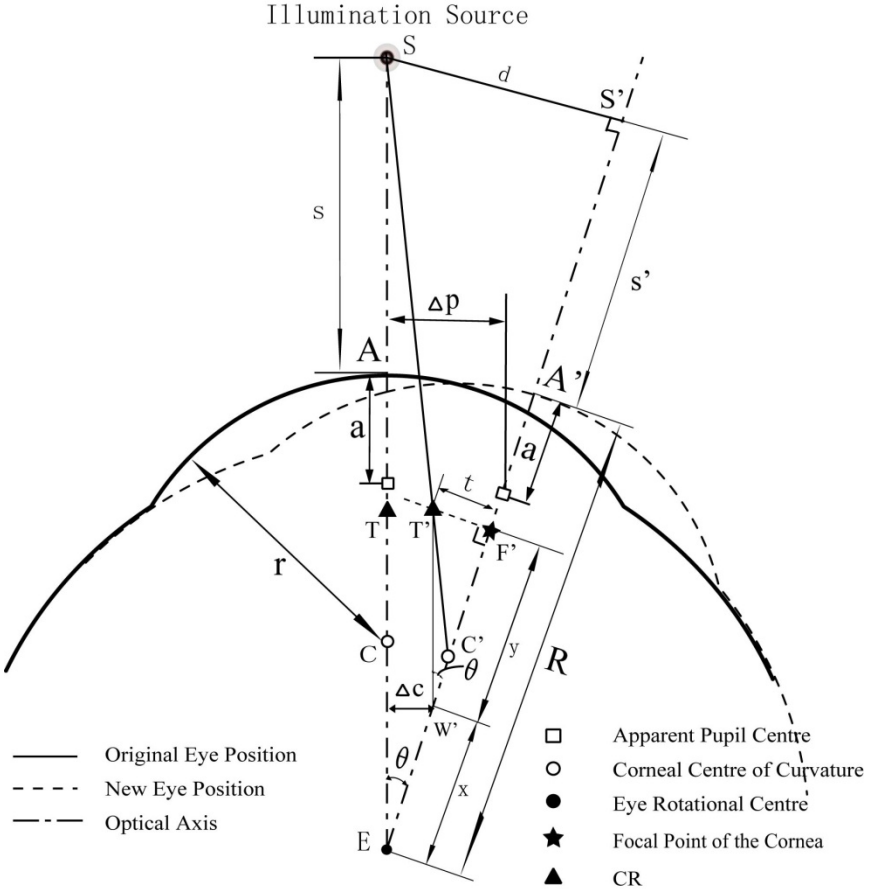

Figure 54. Eye before (solid) and after (dashed) a rotational eye movement when the illumination source is close to the eye.

Because the pupil center is still located on the optical axis of the eye in the new eye

position (see Section 3.2), the pupil offset when the eye rotates can be described by the same equation as Equation 7-2 :

$$\Delta p = (R - a)\sin\theta \tag{7-5}$$

In the new eye position, the illumination source becomes an off-axis object, and its object distance is

$$s' = (s + R)\cos\theta - R \tag{7-6}$$

Note in Figure 54, that S' is the projection of the illumination source on the optical axis of the eye after the eye rotates, and the vector $\overline{SS'} = d$ is computed by using the trigonometric relationship in the right-angled triangle SES′:

$$d = (s + R)\sin\theta \tag{7-7}$$

Similarly, the projection of the CR on the optical axis of the eye lies on the focal point (F′) of the cornea. The right-angled triangles T′C′F′ and S C′S′ are similar, so we get:

$$\frac{t}{d} = \frac{C'F'}{C'S'} = \frac{r/2}{r + s'} \Rightarrow t = \frac{rd}{2(r + s')} \tag{7-8}$$

Based on the triangle T′W′F′, we get:

$$y = \overline{W'F'} = \frac{t}{\tan\theta} = \frac{rd}{2(r + s')\tan\theta} \tag{7-9}$$

The vector $\overline{EW'} = \overline{EF'} - \overline{W'F'}$, so:

$$x = \overline{EW'} = \overline{EF'} - y = (R - r/2) - y = (R - r/2) - \frac{rd}{2(r + s')\tan\theta} \tag{7-10}$$

The displacement of the CR in an eye image is equal to:

$$\Delta c = x\sin\theta = \left[(R - r/2) - \frac{rd}{2(r + s')\tan\theta}\right]\sin\theta \tag{7-11}$$

Inputting Equations 7-6 and 7-7 into Equation 7-11, we get:

$$\Delta c = \left[(R - r/2) - \frac{r(s+R)\sin\theta}{2[r + (s+R)\cos\theta - R]\tan\theta} \right] \sin\theta \tag{7-12}$$

For simplification, we assume small rotational eye movements for which θ is considered as within ±10˚. Thus, we have the following approximations (error introduced is 1% or less if θ is 10˚ or less):

$$\begin{aligned} \sin\theta &\approx \theta \\ \tan\theta &\approx \theta \\ \cos\theta &\approx 1 \end{aligned} \tag{7-13}$$

Inputting these equations into Equation 7-12, we get:

$$\Delta c \approx \left[(R - r/2) - \frac{r(s+R)\theta}{2[r + (s+R) - R]\theta} \right] \theta \tag{7-14}$$

Finally, the rotational gain for any position of an illuminator is equal to:

$$g_{rot} = \frac{\Delta c}{\Delta p} \approx \frac{\left[(R - r/2) - \frac{r(s+R)}{2(s+r)} \right]\theta}{(R-a)\theta} = \frac{R - r/2}{R-a} - \frac{r(s+R)}{2(R-a)(s+r)} \tag{7-15}$$

Inputting values from Table 3 into Equation 7-15, the functional relationship between the rotational gain and the position of the illumination source is represented as:

$$g_{rot} = 0.92 - \frac{0.37(s+13.5)}{s+7.8} \tag{7-16}$$

where input values are in millimeters.

When s is equal to 30mm and 40mm (approximate limits of s for the RIT Eye Tracker), g_{rot} is equal to 0.49 and 0.51, respectively. The plot in Figure 55 graphically shows the functional relationship between the rotational gain and illumination source position described in Equation 7-16. As we can see, as the illumination source is moved further away from the eye, the rotational gain approaches 0.55, which is the value we get

for collimated illumination. Therefore, as with the translational gain, the rotational gain in the collimated illumination condition is a special case of that in the near-source condition.

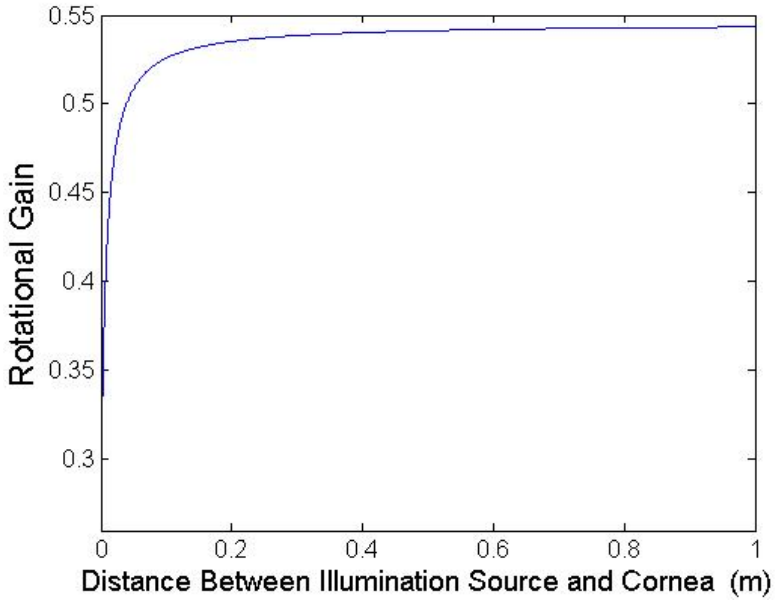

Figure 55. Rotational gain versus position of the illumination source (with respect to the eye).

7.4 Gain Values Validation

Considerable variation exists between individuals in terms of surface curvatures, component separations and axial length (~10%) (Charman, 1994). The above optical derivations based on the Gullstrand simplified eye model were experimentally validated for five observers using the RIT Eye Tracker (Kolakowski & Pelz, 2006 and Li et al., 2008). In the rotational gain measurement, observers were asked to keep their heads and the eye tracking headgear stationary while looking through the calibration points. In the translational gain measurement, observers were asked to fixate on the center calibration

point and, while keeping their eyes and heads still, move the eye tracker's headgear. The headgear was shifted horizontally and vertically over a range of approximately 18 millimeters. A 25-second segment of data from one observer in the trials of the rotational and translational gain measurements are shown in Figure 56. A linear regression was performed on the pupil and CR horizontal data to determine the gain values. The data of Figure 56 is shown in Figure 57 with the best-fit line. The slope of this line is taken as the observer's rotational or translational gain. Both gains, determined for all observers and include the gains derived from the optical relationship of pupil and CR offsets, are shown in Table 7.

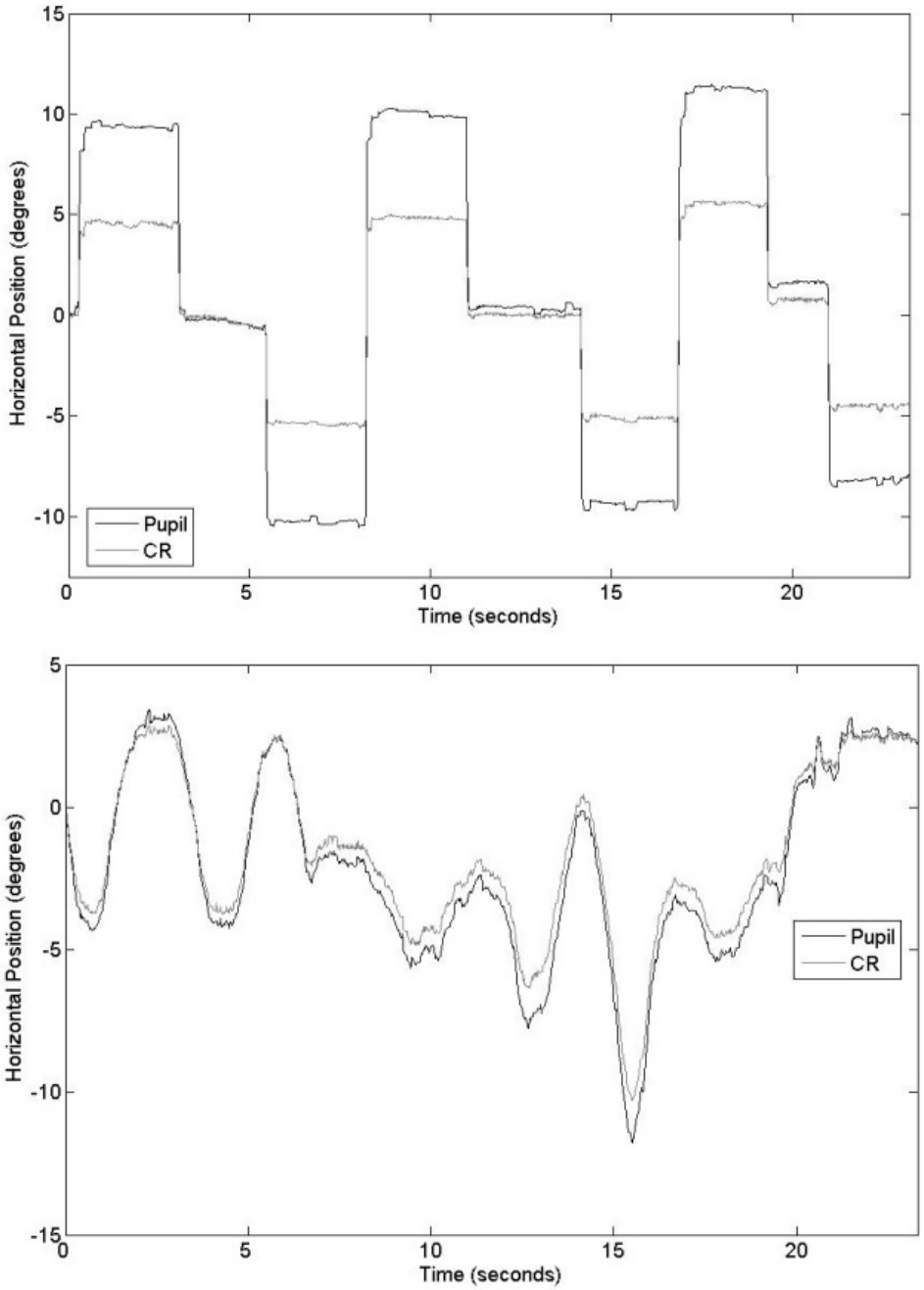

Figure 56. Example data (observer 4) for measuring the rotational gain (top) and translational gain (bottom). Data provided by Kolakowski & Pelz (2006).

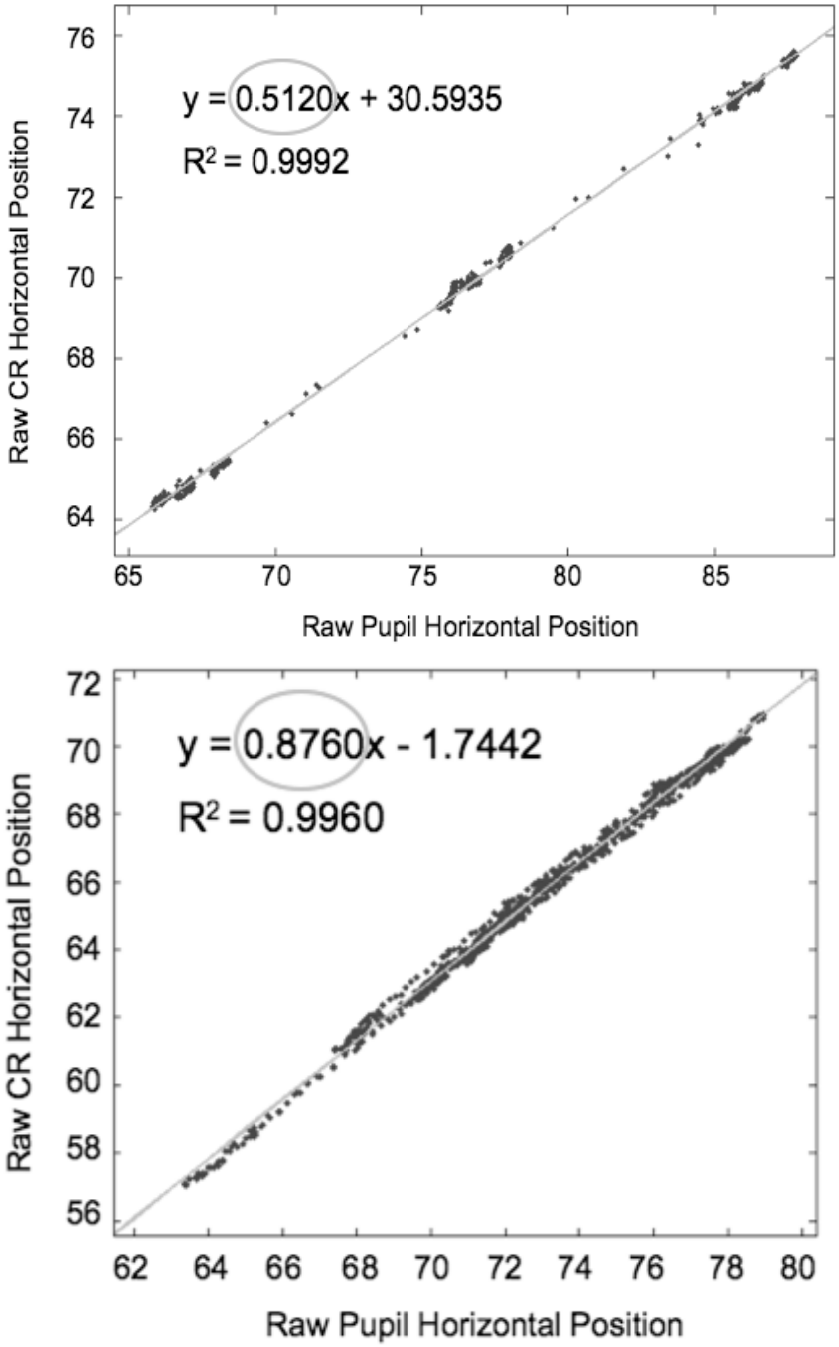

Figure 57. Linear regression for data shown in Figure 56. The overlaid line is the calculated best-fit using least-squares regression, and the slope (circled) of this line is equal to this observer's rotational (top) or translational (bottom) gain. Data provided by Kolakowski & Pelz (2006).

Table 7. Comparison of gain values from experimental measure* and optical derivation

Experimental Measure*			Optical Derivation		
Observer	**Rotational Gain**	**Translational Gain**	**Distance of Source to the Cornea**	**Rotational Gain**	**Translational Gain**
1	0.5161	0.8178			
2	0.5035	0.8653			
3	0.5456	0.8876			
4	0.5120	0.8760			
5	0.4798	0.8720			
Mean	**0.5114**	**0.8637**	**30mm**	**0.49**	**0.90**
SE	0.0106	0.0120	**40mm**	**0.51**	**0.92**

*Experimental measurement data from Kolakowski & Pelz (2006).

The rotational gains calculated in Section 7.3.2 ranged from 0.49 with the IRED 30mm away to 0.51 with the IRED 40mm away from an eye with a corneal radius of 7.8mm. As shown in Table 7, the rotational gain values measured from the five observers ranged from 0.48 to 0.55 with a mean of 0.51 and standard error of 0.01. This mean value matches that derived in Section 7.3.2. The variation is most likely due to variations in the radii of observers' corneas (Charman, 1994), as well as the distance of the IRED to each observer's eye. Therefore, despite the assumptions and simplifications made in deriving these values in Section 7.3, the derived values do closely approximate those determined

experimentally.

The translational gain values derived in Section 7.3.1 ranged from 0.90 with the IRED 30mm away to 0.92 with the IRED 40mm away from the observer's eye. The values calculated for five observers ranged from 0.82 to 0.89 with a mean of 0.86 and standard error of 0.01. This mean value is less than the translational gain values derived theoretically. One reason for the lower gain value may be the intrusion of small rotational eye movements during the task. Even when observers attempt to hold the eye as still as possible, there are microsaccades, tremor and drift (Carpenter, 1988) that will reduce the overall gain. Observers may have also made some rotational eye movements within the fixation point. Because rotational eye movements produce a significantly smaller gain value (about 0.51 for these five observers), this likely accounts for the lower observed translational gain values. A combination of rotational and translational movements will result in an overall gain value (ratio of CR offset to pupil offset) between the rotational and translational gain values derived. The effect was particularly noticeable in data from observer #1 (Figure 58).

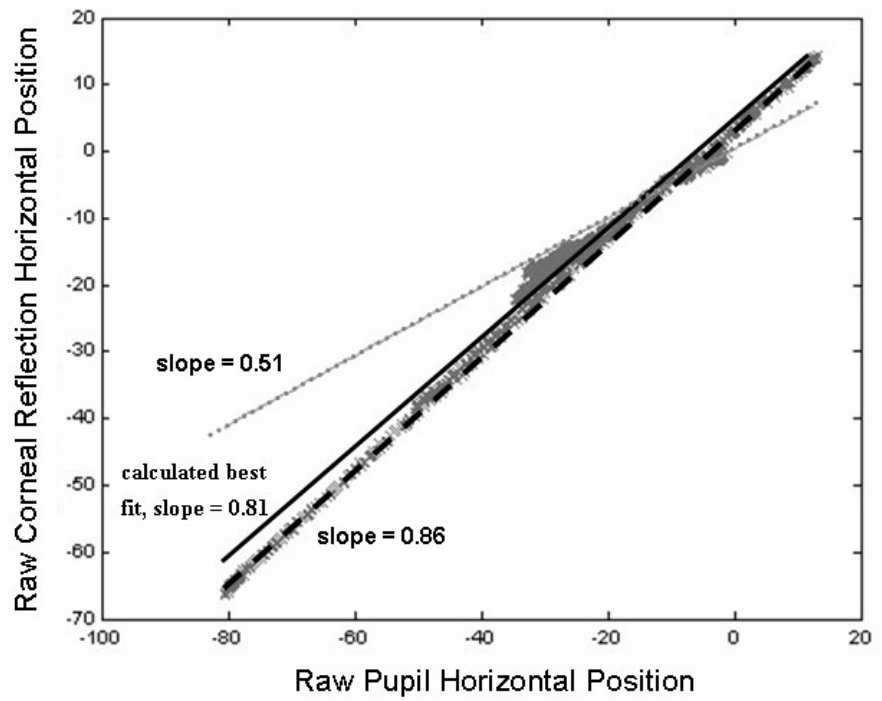

Figure 58. Data for Observer #1 for measuring the translational gain shows the effect of possible coupling of eye movements with rotational movements. Using the average gain over five observers provides a better fit (dashed line) than that calculated for this observer (solid line). The subsection of data around a raw pupil position of -20 appears to fall along a line with slope equal to the average rotational gain of the five observers (dotted line). The data were provided by Kolakowski & Pelz (2006).

Chapter 8 Eye Tracking Optimizations: Structured Illumination and Eye Feature Detection

In this section, we use the rotational gain derived in Chapter 7 to differentiate genuine CRs from false CRs in eye tracking video frames and improve the robustness of our new eye-tracking system prototype. In the new design, we address two limitations of current video-based eye tracking systems: 1) loss of CR during extreme eye movements (large angular rotations), and 2) direct and indirect reflections of ambient light on the cornea being mistaken for the desired CR. Towards solving both of these problems, we use an array of illuminators (3×3 IREDs), what is termed 'structured illumination,' to illuminate the eye. The IREDs, mounted on a piece of supportive Plexiglas in front of the eye, produce multiple CRs in eye images to ensure that – even with extremely eccentric gazes – we can obtain at least one trackable CR (a CR superimposed on either the pupil or iris). This method avoids track losses caused by a single CR "rolling off" the cornea and onto the irregular sclera due to very large eye movements. Thus, the measurement range of the system is improved dramatically. Portions of this chapter have been publisehd in Li et al. (2007) and Li et al. (2008).

8.1 Background

In existing laboratory and commercial video-based eye trackers, track losses caused by extreme eye movements and false detection of the CR due to uncontrolled ambient illumination are significant problems. When the eye rotates to an extreme position, the CR may "roll off" the spherical region of the cornea (Figure 59a) or appear on the eye where it has a low contrast with its surroundings. In this situation, CR detection becomes a non-trivial task and track loss often occurs. As such, video-based systems usually restrain measurement ranges to eye movements within approximately ±25° horizontally and ±20° vertically. Another challenge may occur when more than one specular reflection is a geometrically satisfactory CR (Figure 59b). Under this circumstance, the bright spot closest to the pupil may be taken as the CR (e.g., ASL 6000 series, 2005); this compromise makes the system run smoothly but inevitably introduces error.

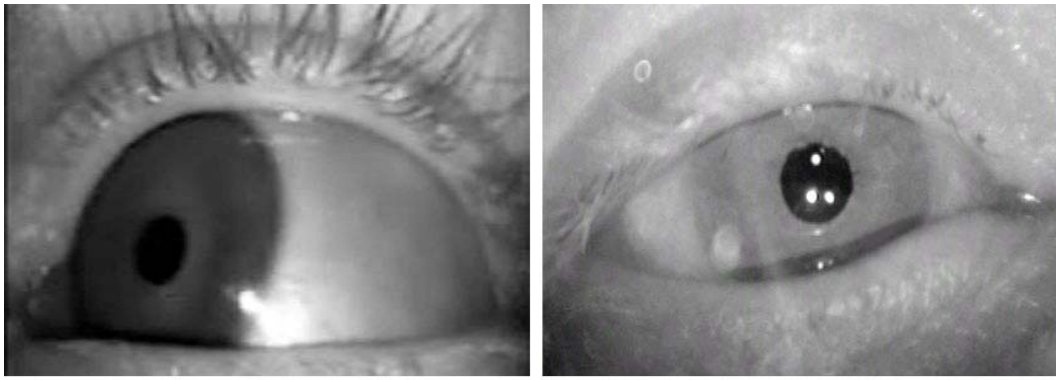

Figure 59. Challenging eye images. (a) A CR rolls over to the aspherical part of the cornea; (b) Spurious specular reflections from metal frames in a lift truck (the bright spot below and to the left of the pupil center is the desired CR). Image (b) from D. Giguère at the IRSST - Safety-Ergonomics Research Program, Canada.

"S*tructured illumination*" utilizes an array of infrared emitting diodes (IREDs) to produce multiple CRs. This configuration ensures that at least one CR is trackable – reflected from the spherical surface of the cornea and superimposed on either the pupil or iris such that it has a high contrast with its surroundings – even with extremely eccentric gazes. One approach for using multiple illuminators in video-based eye tracking has been reported in Haro, Flickner, et al. (2000) and Morimoto, Koons, et al. (2000): two sets of IREDs are mounted on-axis and off-axis with the eye camera and flash alternately to produce bright-pupil and dark-pupil images, and their difference image is used to improve pupil detection. This use of multiple sources is different from the Structured Illumination configuration in the sense that it does not aim to solve the problem of track loss or limited measurement range. A commercial eye tracking system Vision 2000 by El-mar, Inc. (El-Mar) has utilized two IREDs to extend the measurement range of the system (up to horizontally ±40° and vertically ±30°, see Section 5.5.2). However, no technique detail has been made available. Another approach, using a symmetric arrangement of four IREDs around the camera's optical axis to extend the tracking range, has been proposed recently (Hua, Krishnaswamy, et al., 2006); in this approach, at least three out of four CRs need to be detectable for effective tracking. In the approach proposed here, the challenge of spurious specular reflections from external sources or scleral reflections is solved by a two-stage processing approach. First, global and local statistical information from an eye image are used to isolate and determine the pupil center and potential CRs. Second, genuine CRs are separated from false CRs by a CR prediction technique based on the rotational gain introduced in Section 7.3.2.

The rest of this chapter first introduces the prototype implementation of the Structured Illumination configuration in the RIT Lightweight Eye Tracker (Section 8.2).

Detection of potential CRs, the pupil and blinks using statistical information of an eye image are described in Sections 8.3, 8.4 and 8.5, respectively. The application of using rotational gain to discriminate genuine CRs is presented in Section 8.6.

8.2 Illumination Configuration

The new illumination configuration uses nine IREDs in a 3×3 array as opposed to only that many eye tracking systems use. The illumination sources are mounted equally spaced on a transparent supporting frame with an eye camera placed at the lower-right of the center IRED (Figure 60).

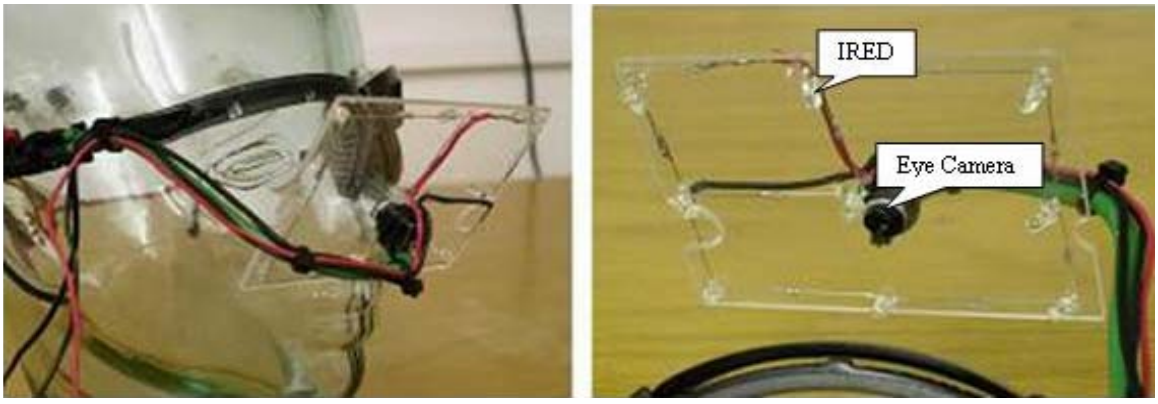

Figure 60. Prototype of the Structured Illumination

Figure 61 shows two example images with multiple CRs from the new design. Note that reflections from some of the illuminators are visible even during extreme eye movements (Figure 61b). However, many artifacts are observed in the eye images. Actually more artifacts may be seen in daylight, due to spurious reflections from multiple IREDs and/or extraneous ambient illumination. The aspherical nature of the

cornea-scleral boundary is another source of irregular reflections.

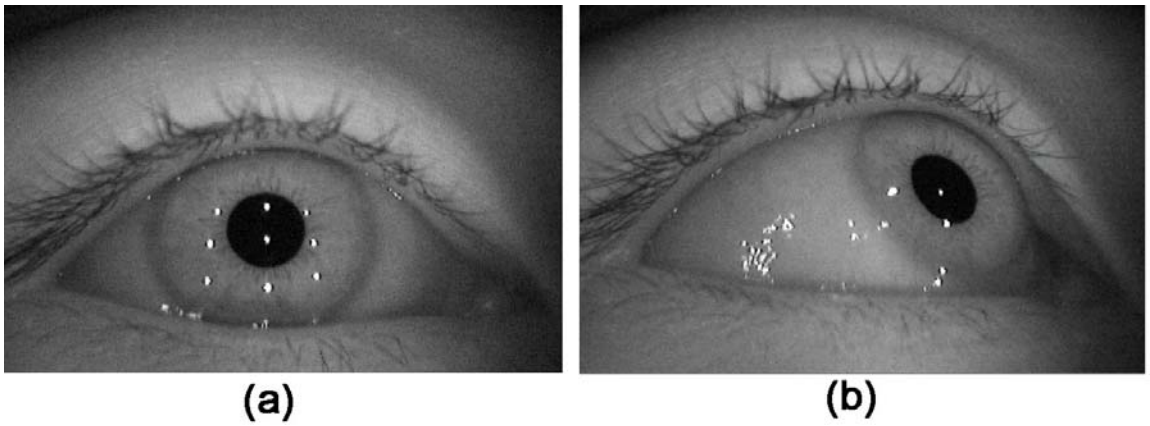

Figure 61. Eye images illuminated by the Structured Illumination. (a) The eye looking straight ahead; (b) the eye moved to an extreme position.

Based on our testing, at least three CRs could be detected in each video frame when eye movements up to ±40° horizontally and ± 25° vertically were recorded.

8.3 Potential CR Detection

Detection of potential CRs consists of several steps. A flowchart is shown in Figure 62. An eye image is first smoothed by a 5×5 Gaussian filter with a standard deviation of 2 pixels to reduce noise effects. The top 1% of the brightest pixels in the image are then selected; CR pixels are among these brightest pixels. After region-labeling, connected components larger than 0.15% of the image size are eliminated (Figure 63a and b). Additional spurious specular reflections are removed by a local contrast technique, which only accepts potential CRs that have high contrast with a small set of their neighboring pixels. In other words, for each connected component (with pixel values in the top 1% of

pixel values in the image), the grayscale difference between its mean and that of its surroundings must be larger than a threshold (50 in our implementation) to be selected as a *potential CR* (Figure 63c and d). The surroundings are defined as eight neighboring rectangular regions, each with width and height equal to the width and height of the connected component. Using the local contrast technique, a significant amount of spurious CRs is eliminated. However, several artifacts are still confused with the genuine CRs.

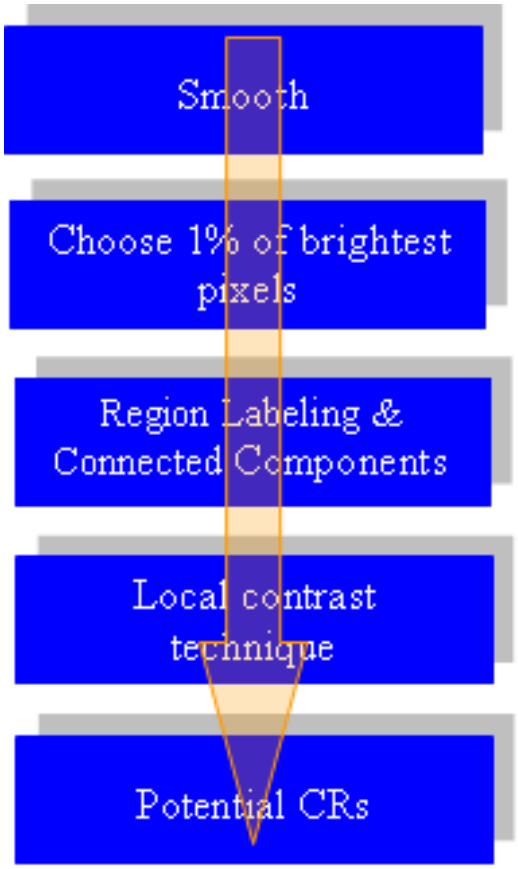

Figure 62. Flowchart of the potential CR detection

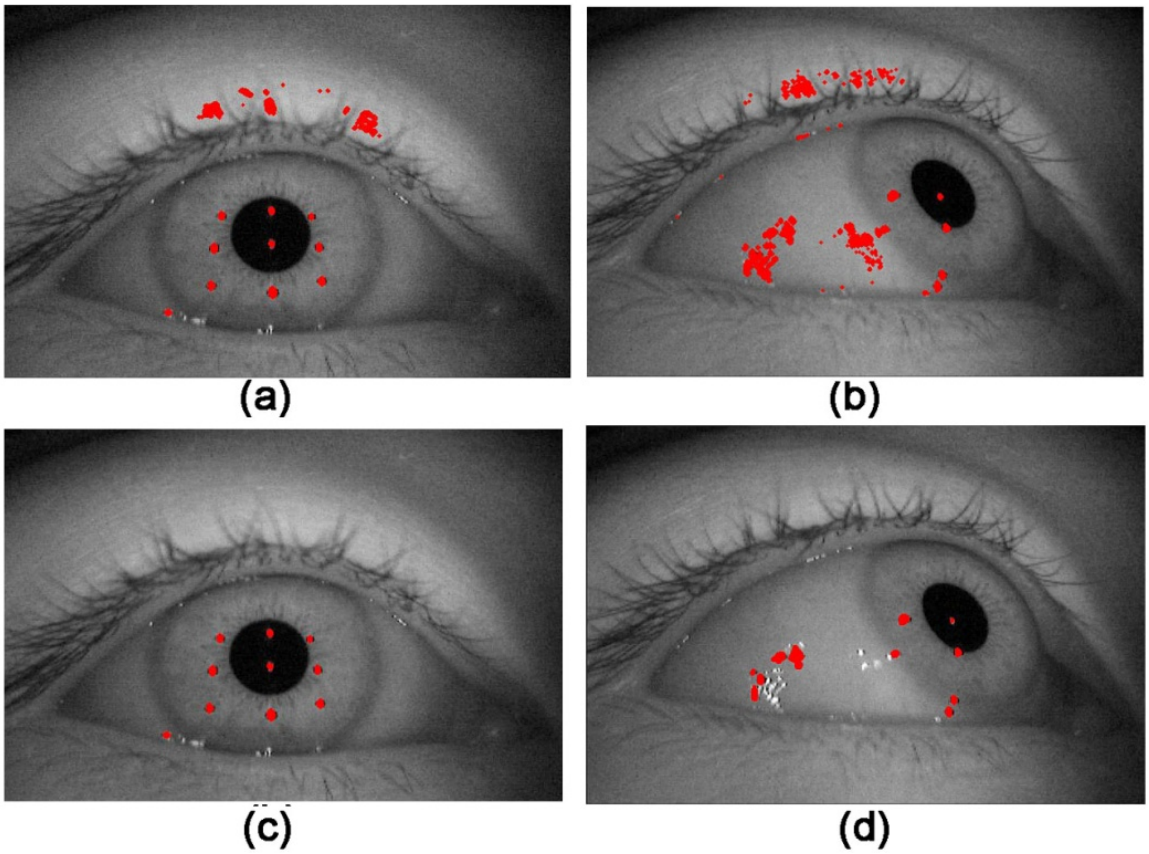

Figure 63. Eye images in CR detection. (a)(b) Connected components with pixel values in top 1% (highlighted in red color) after eliminating components larger than 0.15% of image size; (c)(d) Potential CRs after the local contrast technique is applied.

8.4 Pupil Detection

The basic steps for detecting the pupil are shown in the flowchart in Figure 64. In order to detect the pupil in an eye image, the potential CRs are first removed and their remaining holes are filled with the mean gray values of their surrounding areas (Figure 65 a and b).

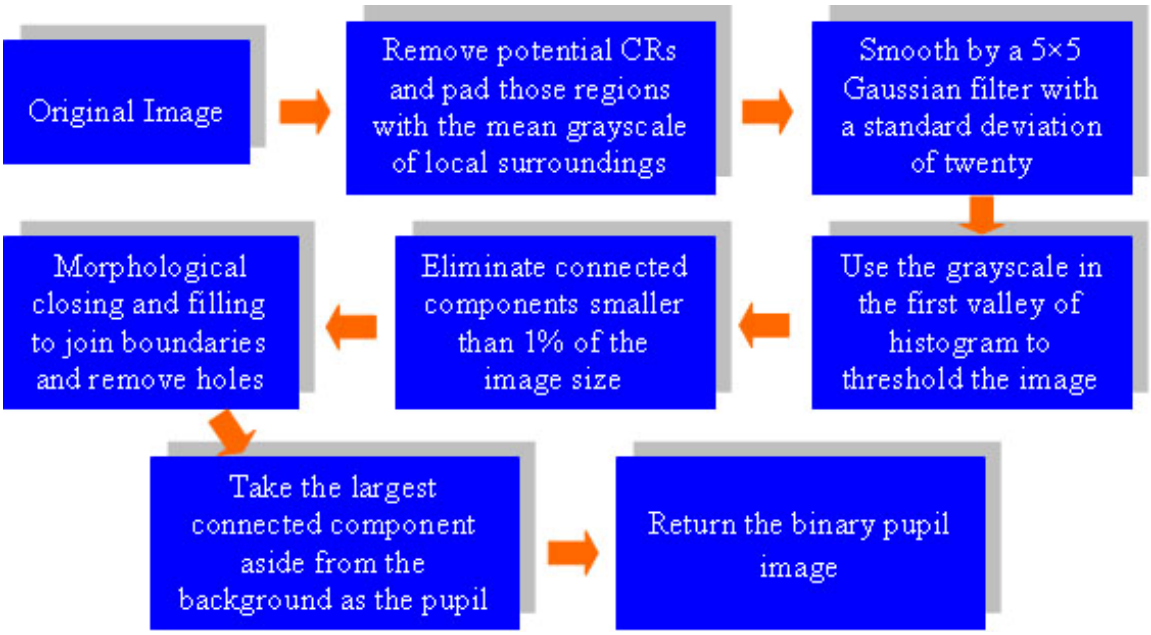

Figure 64. Flowchart of the Pupil detection

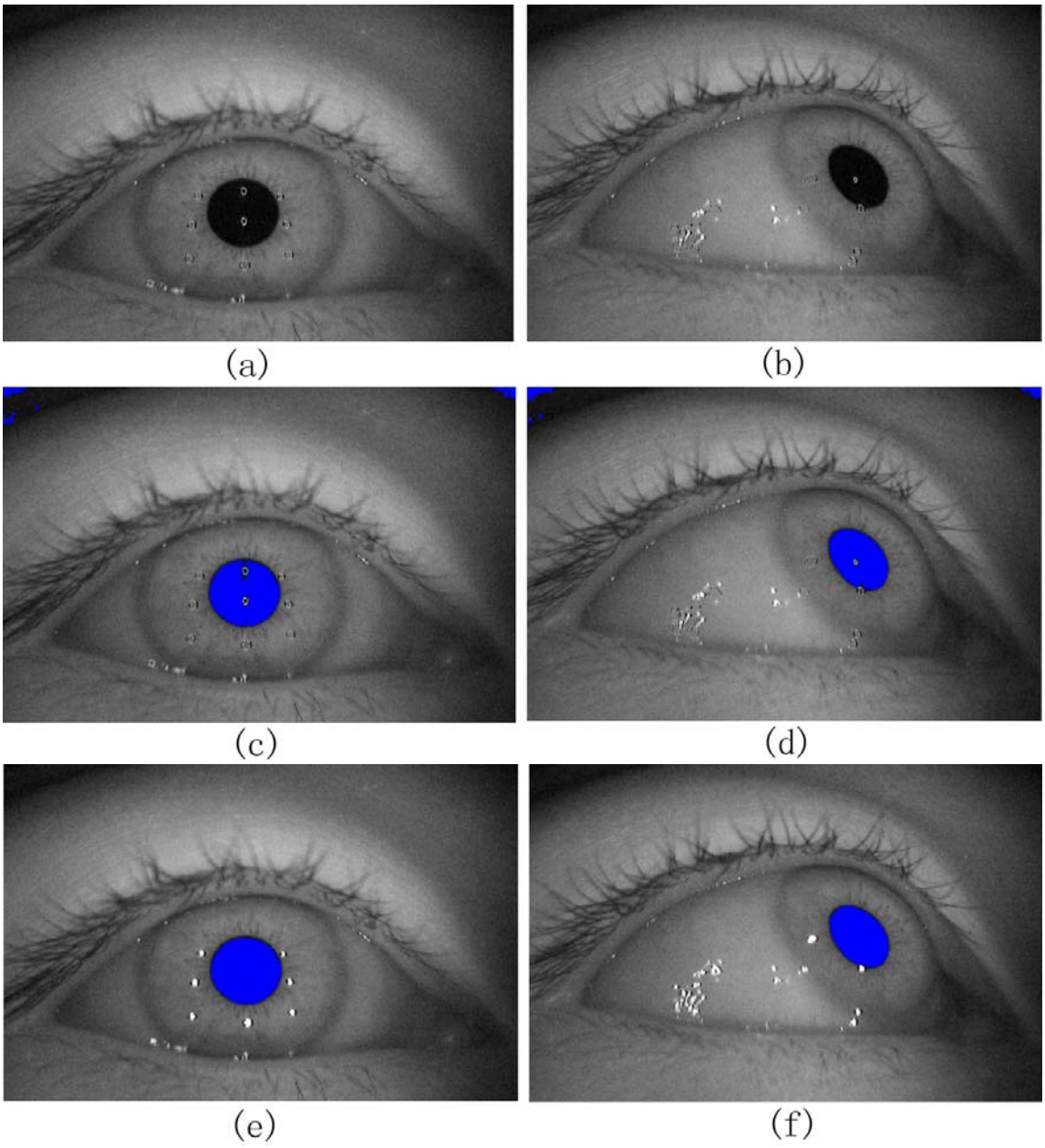

Figure 65. Pupil detection

The eye images after CR removal are blurred by a Gaussian filter with a large

standard deviation of 20 pixels. The size of the image is 720×480. This step aims to smooth sharp edges caused by padding CR holes and to reduce the influence of small occlusions (such as those due to eyelashes). The first valley in the histogram of the blurred image (Figure 66) represents the pupil grayscale limit (where the pupil is a dark disk in the image); a threshold operation is thus applied to select the pupil pixels as those whose values fall below this grayscale value.

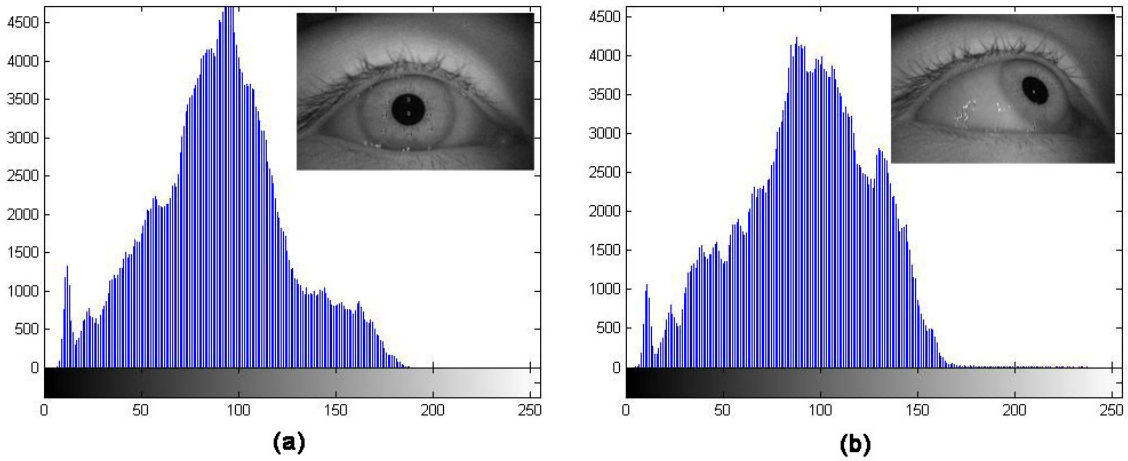

Figure 66. Use of the grayscale in the first valley of the histogram to threshold the eye image

Connected components smaller than 1% of the image size are eliminated (Figure 65 c and d). Note that some artifacts in the upper corners and holes inside the pupil area can still be seen. Morphological closing and filling operations are then used to join boundaries and remove holes inside all connected components. The pupil detection process is finalized by choosing the largest connected component in the binary image, excluding the background (Figure 65 e and f).

8.5 Blink Detection

The detection of blinks is based on the following calculation:

$$\text{circularity} = 4\,\pi\,(\text{area}/\text{perimeter}^{\wedge 2}) \qquad (8\text{-}1)$$

The circularity evaluates how close a shape is to a perfect circle (circularity = 1). If the eye camera is positioned in front of the eye (as in our case), the pupil will appear as a circle or ellipse seen in the image during eye movements. When a blink happens, the circularity of the detected "pupil" region will be significantly less than unity. In our implementation, if the circularity of the detected "pupil" region is below 0.7, the program considers that a blink is detected and a text message is triggered (Figure 67). Note as shown in Figure 67c, when the pupil is partially occluded, the frame is still considered in a blink because the detected pupil arc is incomplete. As a result, the circularity calculated from the fitted ellipse is below the predefined value (Figure 67d). For cases like this, other techniques might be needed to exclude the irregular edge due to occlusion and rely only on the visible circular arc in the fitting process.

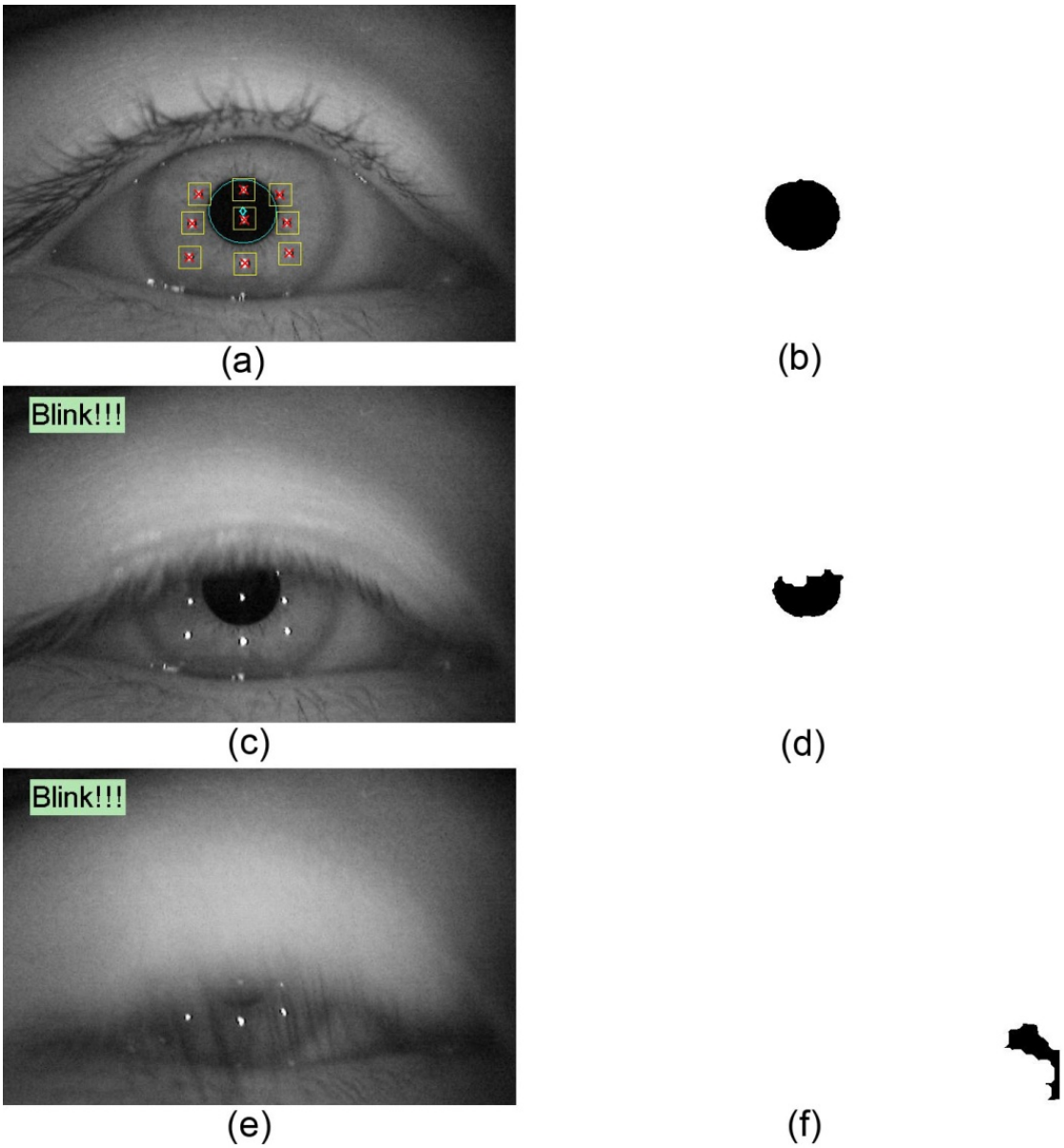

Figure 67. Eye images during a blink. (a) Before the blink – circularity = 0.94; (c) During a blink – circularity = 0.58; (e) At the end of a blink – circularity = 0.26. Images (b) (d) and (f) are their corresponding binary images during pupil detection.

8.6 CR Prediction

In this section, we use the rotational gain derived in Section 7.3.2 to differentiate desired CRs (actual reflections of the eye tracker's illumination sources) from potential CRs. As noted in from Sections 7.2.2 and 7.3.2, the CR moves only about half the distance as the pupil center when the eye rotates for collimated or near-source illumination conditions. The rotational gain (g_{rot}) describes the relationship between the CR (Δc) and the pupil (Δp) offsets:

$$g_{rot} = \frac{\Delta c}{\Delta p} \approx 0.5 \tag{8-2}$$

After rearranging and expanding Equation 8-2, we obtain:

$$CR_{current} - CR_{previous} = \Delta c \approx 0.5\Delta p \tag{8-3}$$

$$CR_{current} \approx 0.5\Delta p + CR_{previous} \tag{8-4}$$

This last equation shows that a displaced CR position can be approximately predicted if the pupil offset between the current and previous frame and the CR position in the previous frame are known.

In our implementation, a user selects nine CR positions in the first frame. Nine predicted CR locations for each following frame are then calculated using Equation 8-4. When a potential CR falls inside a predefined square (length of side = 30 pixels) centered on a predicted CR location, it is identified as a genuine CR. This square region tolerates a

small amount of translational eye movement (e.g., eye camera movement with respect to the head) and, as well, individual variations between observers, such that the genuine CRs will still fit within these regions. If a potential CR does not fall inside this square, it is considered to be an artifact and excluded from the final genuine CR list (Figure 68). If multiple potential CRs fall in a predefined prediction region, which is very rarely the case, the one closest to the center of the prediction square is taken as the genuine CR. This simplification may introduce errors, if the 'detected' CR is not the real one. However, the error will be minor (unlike the case discussed in Section 6.1 c), as the positional difference between this mistaken CR and genuine CR is small. The circumferences and centers of the pupil and CRs are calculated by a direct least-squares ellipse-fitting technique introduced in Fitzgibbon, Pilu et al. (1999).

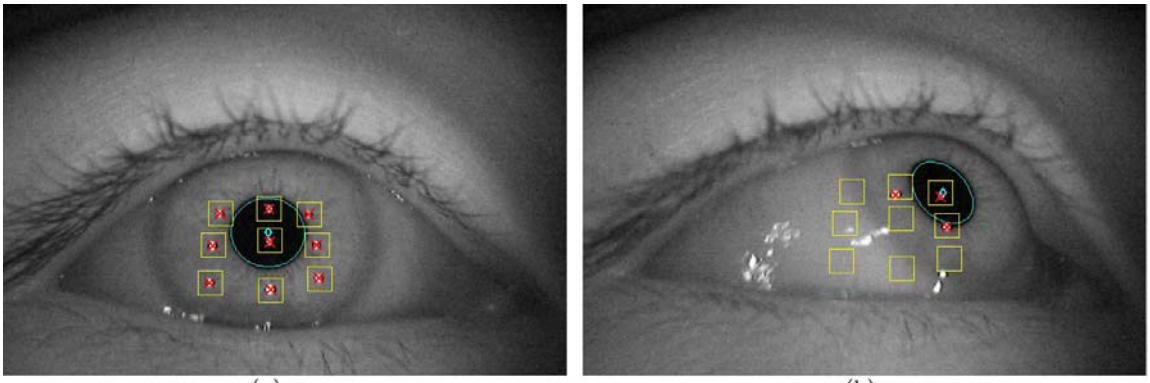

Figure 68. Eye images using CR predictions by the pupil-CR offset ratio. (a)The eye looking straight ahead; (b) The eye moved to an extreme position. Yellow squares represent predicted CR locations, red crosses represent detected genuine CRs, the cyan ellipse represents the pupil boundary, and the cyan diamond represents the pupil centroid.

The CR prediction technique robustly distinguishes genuine CRs even when there are spurious specular reflections in the eye image. When a CR, such as the one in Figure 61b, falls close to the limbus (boundary between the iris and sclera), the prediction for

this CR may fail. If this happens, there should still be other detectable CRs in the eye image, which can be used to estimate the POR vector. Another limitation occurs when the brightness of a CR, due to motion smear during a rapid eye movement, decreases below a significant level (upper-right CR in Figure 69). An ideal solution would use a high-speed camera, as discussed in Section 6.2, to capture sharp eye images during a fast eye movement. If a high-speed camera is not an option, our eye tracking system using structure illumination is still able to rely on those CRs that can be identified successfully. Furthermore, the effects of motion blur can be minimized by averaging (or median filtering) the P-CR vectors derived from multiple CR points.

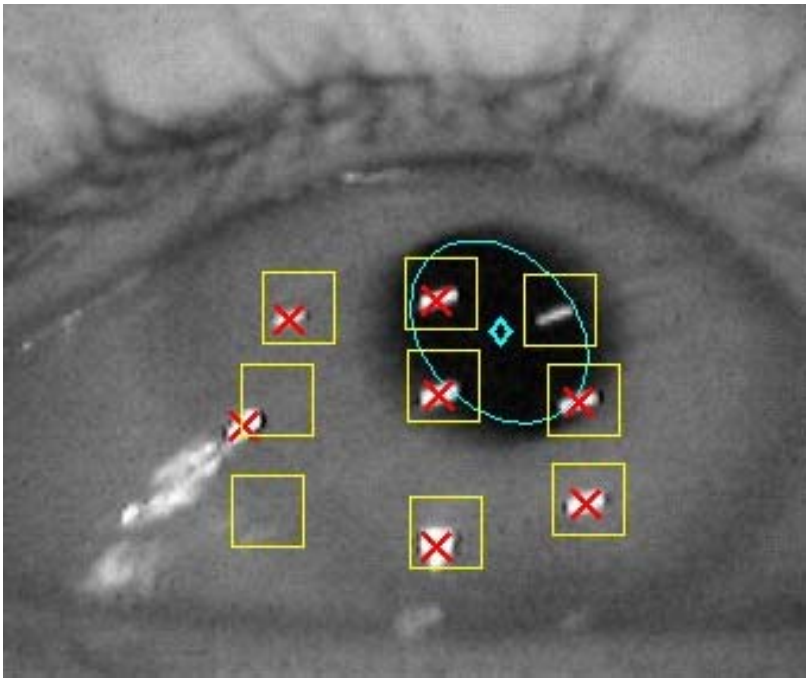

Figure 69. CR prediction for a blurred eye image (due to a rapid eye movement). Note that the upper-right CR has not been detected.

Beside the application in predicting the CR position to eliminate artifacts due to spurious reflections, the translational and rotational gains derived in Chapter 7 have also been used to effectively compensate for camera movement and permit noise reduction without degradation in temporal resolution outputs (Kolakowski & Pelz, 2006; Li, Kolakowski, et al., 2008).

In the technique, the CR data are only used to determine translational movements as shown in the following equation:

$$P_{trans} = \frac{g_{rot} P_{track} - CR_{track}}{g_{rot} - g_{trans}} \; P_{trans} = \frac{g_{rot} P_{track} - CR_{track}}{g_{rot} - g_{trans}} \tag{8-5}$$

where P_{track} and CR_{track} are raw pupil and CR position data from the eye tracker, and P_{trans} represents the amount that the pupil center has moved in the eye image due to translational movements of the camera with respect to the eye. Since the camera moves infrequently and relatively slowly, this vector may be smoothed before being extracted from the tracked pupil position, P_{track}, without adverse effect. This allows for reduction of noise due to CR position estimation without loss of information about rotational eye movements, since P_{trans} contains no information about rotational eye movements, only translational movements between the eye and the camera. The final eye-in-head data – equal to the pupil position vector minus the camera position vector – will have a noise level comparable to that in the less-noisy pupil position data. This method, relying on the rotational and translational gains to differentiate two types of eye movements, does not require any additional processing and can be applied to any eye tracker that tracks the pupil and CR (Li, Kolakowski, & Pelz, 2008).

Chapter 9 Eye Tracking Applications: Smooth Pursuit Dynamics in Controlled Static Conditions

Quantitative measure of the dynamics of eye movements is one of the important applications in eye tracking techniques. The following two chapters describe eye-tracking experiments designed to study the characteristics of smooth-pursuit eye movements under two different conditions. The first one is a controlled condition (Chapter 9), where the head movements of the subjects were constrained by using a chinrest and a headrest. In contrast, the subjects were allowed to freely move their heads (a natural condition) during the eye tracking study in a car in Chapter 10. Another important difference between the two studies is that the visual stimuli were moving at a distance from the subject in the controlled condition where they were placed relatively close to the eye in the unconstrained condition. This chapter provides the details of the application of video-based eye tracking techniques in studying smooth pursuit in the controlled condition.

9.1 Background

In traditional smooth pursuit experiments, small targets (e.g., <1°) against homogenous backgrounds were used to drive smooth pursuit eye movements (e.g., Rashbass, 1961). However, in our everyday lives, we frequently track large and complex targets against richly textured backgrounds. For example, we track an ad on a bus driving by, or are sitting on a sofa while watching car-racing in an action movie. Recently, exploring the performance of human pursuit system in tracking extended targets with or without a complex background has drawn increasing interest. Some researchers (Hashiba, Hattori, et al., 1996; Masson, Proteau et al., 1995; Niemann & Hoffmann, 1997) found that the velocity of smooth pursuit was reduced when subjects tracked a small target against a still, textured background, while pursuit performance was enhanced when the small target and textured background moved in the same direction. Other researchers (Heinen & Watamaniuk, 1998; Pola & Wyatt, 1985; van den Berg & Collewijn, 1986) discovered that pursuit ability was also improved when tracking an extended stimulus. In these studies, however, pursuit performances were studied by using artificial targets such as small targets moving sinusoidally (Pola & Wyatt, 1985) or along a sawtooth trajectory (van den Berg & Collewijn, 1986), mostly with a synthetic background, such as a random dot pattern (Collewijn & Tamminga, 1984; Hashiba, Hattori, et al., 1996; Masson, Proteau, et al., 1995). In addition, most of these studies were conducted with relatively low target velocities (< 20 deg/s).

The limit of pursuit velocity has been studied by using small targets, while the effect of target size on the velocity limit is not well understood. Westheimer (1954) reported that when the target velocity exceeded 30 deg/sec, eye velocity became

substantially slower than the target velocity and saccades began to intrude frequently in the records. A similar result was reported by Robinson (1965), who showed that the eye velocity approached saturation at 20 deg/sec. These two studies suggest that human smooth pursuit is limited to a target velocity of about 30 deg/sec, and since then, this result has been widely quoted.

However, recent studies have reported significantly greater saturation velocities. Meyer, Lasker et al. (1985) claimed that the upper limit of smooth pursuit is up to 87 deg/sec (gain equal to 0.87 for the target velocity of 100 deg/sec) and beyond this limit, the eye velocity is saturated with large variability (Figure 70). Note that in their study, the limit of the target velocity was actually rounded from 90 deg/sec to 100 deg/sec, which resulted in increasing the eye velocity limit from 79 deg/sec to 87 deg/sec. The authors argued that “the difference of 10 deg/sec is small given the variability in the data” (Meyer, Lasker, et al., 1985).

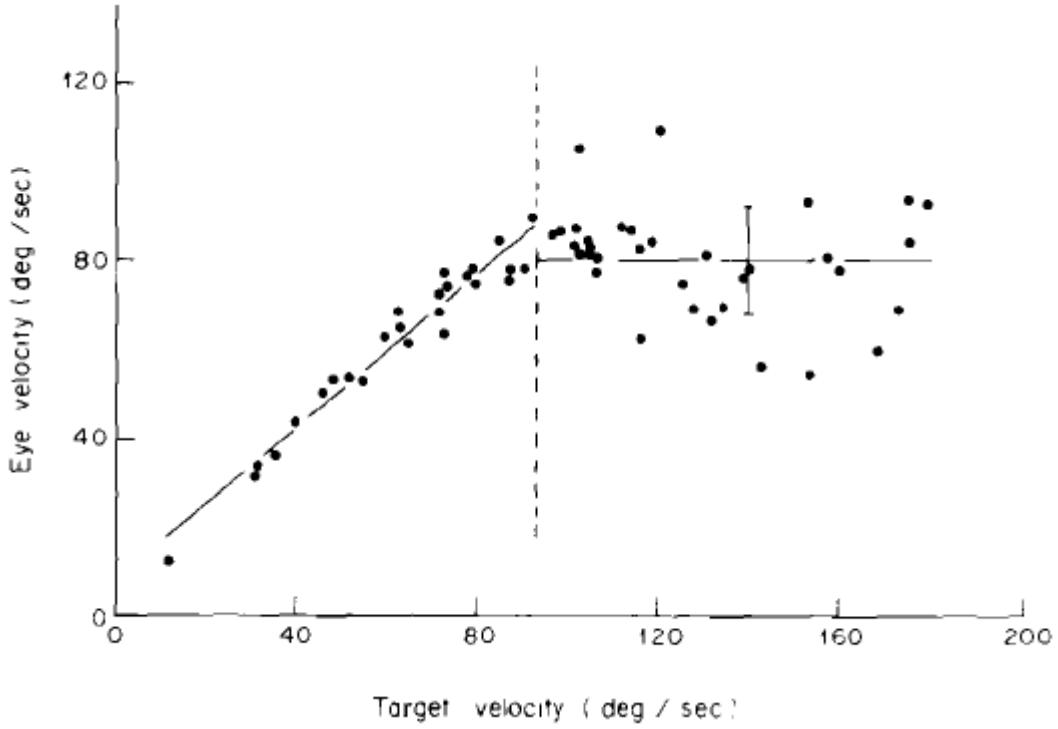

Figure 70. Eye velocity versus target velocity (subject K.C.). At target velocities over 100 deg/sec, eye velocities appear saturated with a large variability. The vertical, solid line at high velocities indicate one SD of velocity variability. Image from Meyer, Lasker, et al. (1985).

In all the studies described to this point, targets underwent "passive motion", i.e., their motion was generated by the experimenter, and the subjects' task was simply to follow the target. It has been reported that significantly greater limits are possible under conditions of "active motion," where subjects try to follow motion paths they generate themselves. Collewijn, Steinman, et al. (1985) reported that the smooth pursuit eye velocity limit was between 50 and 100 deg/sec when the target was moved by the experimenter, but the limit was increased to between 125 and 175 deg/sec when the target motion was generated by the subject. The large range within the passive and active motion was ascribed to individual differences.

In the present study, the effect of stimulus extent on smooth pursuit was studied in two experiments by using realistic images (color photographs of an apple and flower subtending 2° and 17°, respectively) in comparison with a small dot subtending a fraction of a degree. Subjects were allowed to freely pick any detail in the two extended images to track. In Experiment 1, five different target velocities ranging from 16 to 78 deg/sec were used to test the relationship between the smooth pursuit performance and target size when tracking realistic images (along with the influence of target velocity). In Experiment 2, target velocities were extended to 94 and 112 deg/sec to explore the influence of stimulus size on the upper limit of smooth pursuit velocity.

9.2 General Methods

9.2.1 Experimental Setup

Subjects were seated in a comfortable chair at a viewing distance of 337cm. Head movements were stabilized by a chinrest and a headrest (Figure 71). Black foam-core boards were placed on both sides of the subject head to limit the viewing angle to about ±30°. Three first-surface mirrors relayed the stimulus from an LCD projector (Canon LV5300) onto the wall in front of the subject (Figure 72). Displaying the moving images in this way minimized potential artifacts such as motion blur or image 'studder' that would be caused by moving the image on the CRT, LCD, or similar displays, while maintaining a large field of view. Furthermore, the rotating mirrors caused real motion of the stimuli rather than apparent motion of traditional displays. The mirrors were driven by a DC motor to project a moving stimulus. The rotational axis of the motor and the

center of the projector were on the same plane, a distance of 337cm away from the wall showing the stimuli. The stimuli moved horizontally from left to right at velocities controlled by adjusting the voltage to the motor using a constant-current power supply (Hewlett-Packard 6205B).

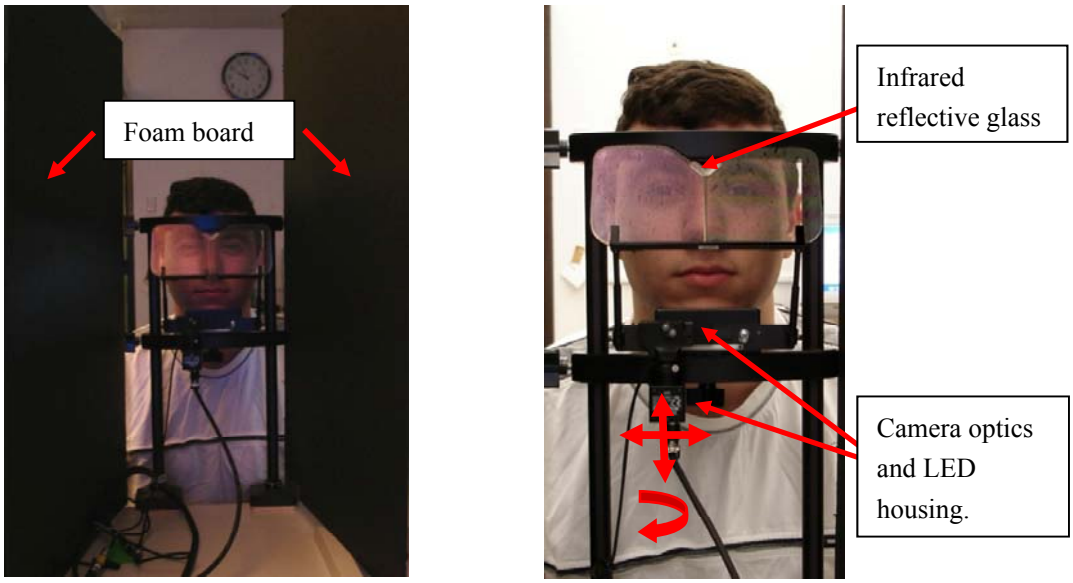

Figure 71. Experimental setup and eye tracker. Two core-foam boards shown in the left picture were used to limit the subject's viewing angle to ±30º. Only the left side of the frontal glass from the reader's view is coated with an infrared reflective coating; the other side is a regular transparent glass. The camera optics has three degrees of freedom of adjustment to center an eye image.

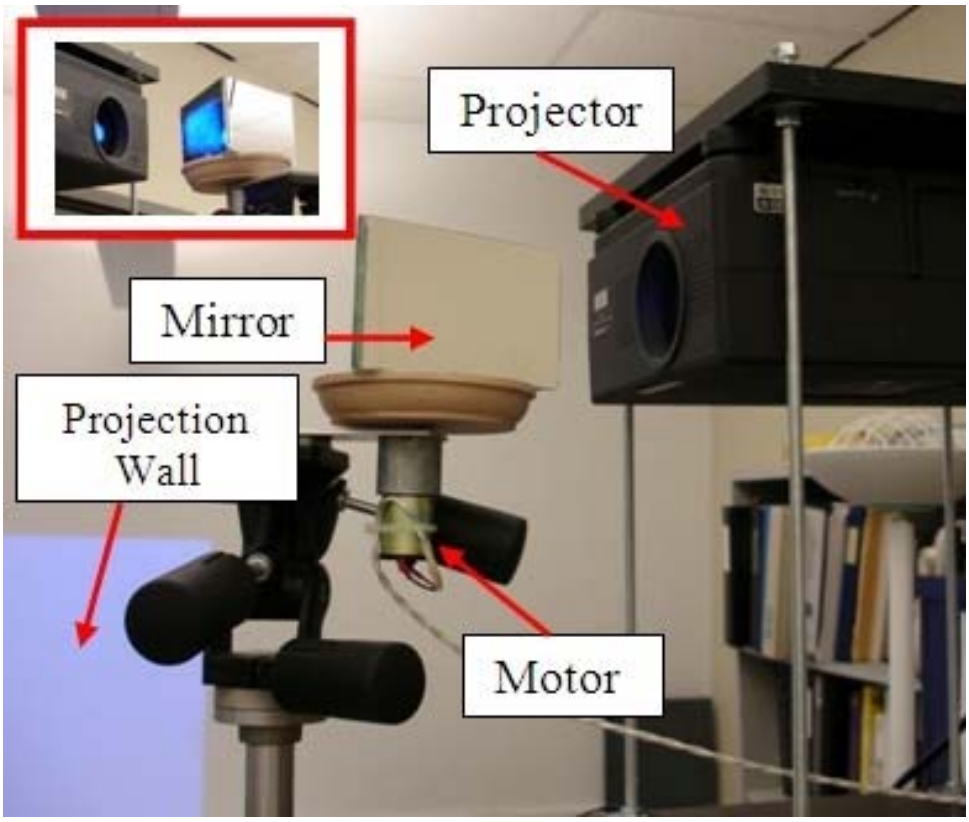

Figure 72. Setup of the projector, DC motor and three first-surface mirrors. Three first-surface mirrors relayed a stimulus image from the projector to the wall. The mirror array sat in a wood supporter which was driven to spin by a DC motor. The small picture on the left-upper corner shows the setup viewed in an opposite direction.

9.2.2 Visual Stimuli

The full-field 1280×768 pixel image subtended about 17×10 degrees of visual angle. The images were a small red dot on a black background, a color apple on a black background, and a color flower image (Figure 73). The dot, apple, and flower images subtended horizontally about 0.1, 2 and 17 degrees of visual angle, respectively. The stimuli moved horizontally from left to right. Stationary photographs were used rather than videos so

that all pixels in the image would move at the same rate, thus avoiding other kinds of motions (e.g., non-rigid motion).

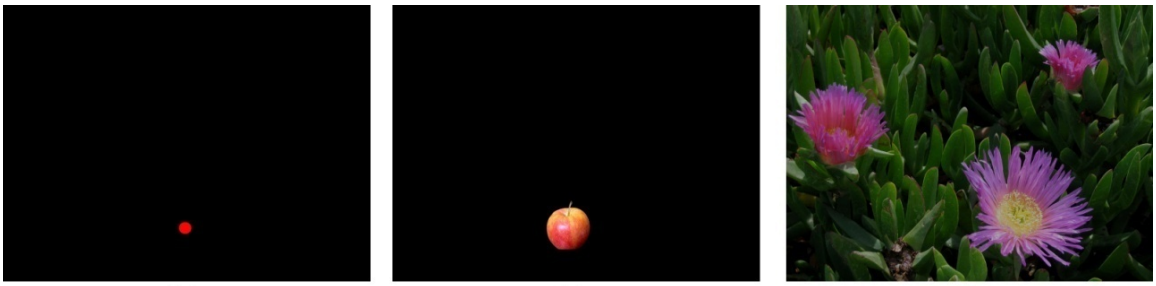

Figure 73. Three stimulus images. (a) The dot (enlarged for display purpose); (b) The apple; (c) The flower image

9.2.3 Eye Movement Recording

A high-speed eye tracker H6-HS by Applied Science Laboratories was used to record eye movements. This is a video-based eye tracker relying on the positional difference between the pupil center and first-surface corneal reflection (CR) to map the line of sight (see Section 5.5). The setup is shown in Figure 71. The optical module of the high-speed camera was mounted on the chinrest, and had three degrees of freedom of adjustment to capture the image of the subject's right eye. A thin-film coating, transparent in the visible spectrum, reflected infrared illumination from an IRED toward the subject's eye and back to the camera. The system recorded monocular eye position at 240 Hz (4.17 msec sample rate). Viewing was binocular. The eye tracker was calibrated by instructing the subject to fixate successively on calibration targets at seventeen known positions on the stimulus plane. The accuracy of the eye tracker within the calibration region is approximately 1º.

9.2.4 Experimental Procedure

Before each experiment, subjects were informed that an image (shown to them in advance) would move from left to right. They were asked to track the stimulus as smoothly as possible, while keeping their eyes waiting to the left when there was no stimulus. Each experiment was divided into three sessions (for the three images). For the apple and flower images, subjects were instructed to pick any point on the image to track, but to be consistent within each trial. In other words, they were free to track any detail of the image. To reduce uncertainty, subjects were told if the target velocity was an increase or decrease from the previous trial. All lights were turned off expect for a dim lamp in a neighboring room.

9.2.5 Eye Tracking Data Analysis

Gaze data were stored on a PC and analyzed "off line". Eye-position data from each trial were plotted and analyzed visually. In the example eye-position trace from Experiment 1 (Figure 74), two thick horizontal lines represent the calibration region (±12°) of the eye tracker, the data within which were considered the most reliable. In Experiment 2, the eye tracker was calibrated within the region between -6° and 18°, because the subject was observed to need more time to catch up with high-velocity targets based on a pilot experiment. Two thin horizontal lines describe the visually coded region (±25°). Any data falling outside these two lines were not included in analysis because these data were considered as having low reliability. The tilted parallel lines indicate the actual target velocity. The steady-state pursuit segment (via visual inspection) within the two vertical lines was linearly fitted, and its slope was calculated as the smooth pursuit velocity of the

trial.

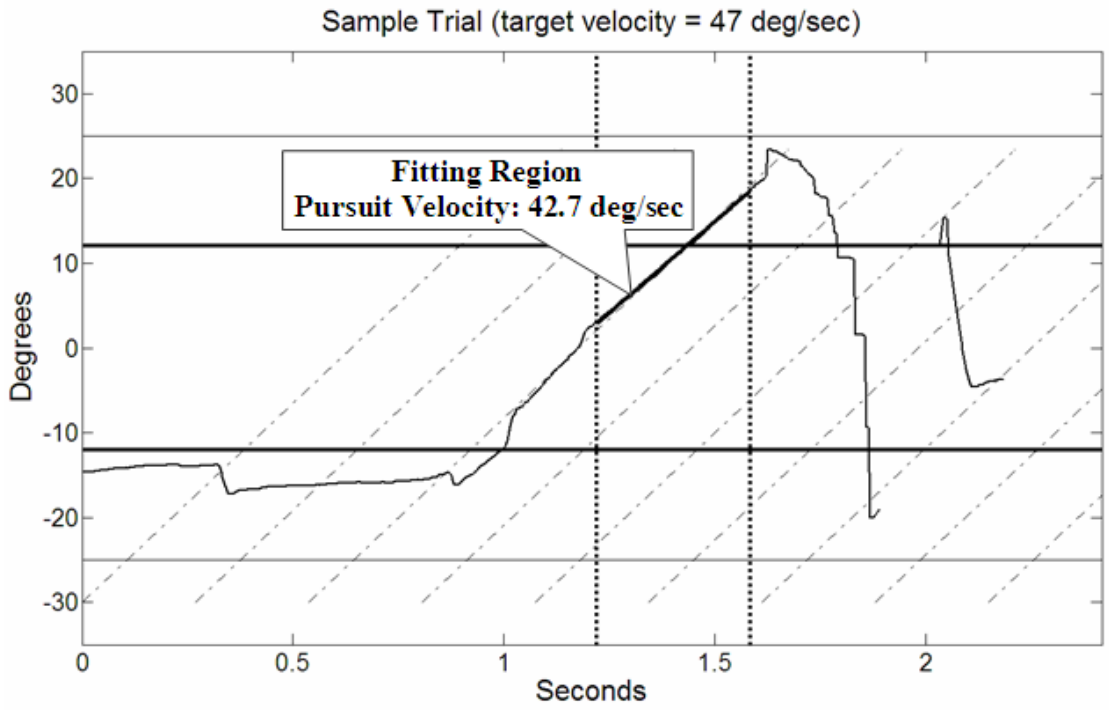

Figure 74. An example eye-position trace in Experiment 1. Two thick horizontal lines represented the calibration region of the eye tracker (the most-reliable region). Two thin horizontal lines described the visually coded region (±25°). Two dashed vertical lines depicted a steady-state pursuit segment within which eye-position samples were linearly fitted and its slope (solid, slanted segment) was calculated as the velocity of the smooth pursuit. The slanted parallel lines represented the real target velocity.

9.3 Experiment 1: Effects of Stimulus Size and Velocity on Smooth Pursuit

9.3.1 Subjects

Thirteen subjects (students and faculty at the Rochester Institute of Technology)

participated in the experiment. All subjects had normal or corrected-to-normal vision. There were two females and eleven males from 21-43 years old (average 27.2). Subjects provided informed consent and received an honorarium for participating. Three subjects had a pupil or CR occluded by their upper eyelids (causing unacceptable amount of tracking loss,) and one subject had a specular reflection of the IRED on the frame of his eyeglasses (mistaken by the eye tracker for the desired CR). The data from these subjects were excluded. Consequently, eye-tracking data from the two females and only seven of the males were used in the analysis.

9.3.2 Data Collection

Subjects were given a short break (1-2 minutes off the chinrest) between each session as the experimenter changed the stimulus. Before and after each session, the calibration of the eye tracker was checked and recalibrated if necessary. Five target velocities were used in this experiment, 16, 31, 47, 63 and 78 deg/sec. At each velocity in each session, 20 trials (belonging to two blocks) for two subjects and ten trials (belonging to one block) for the other seven subjects were collected. As a result, a total of 1650 eye-position traces in 165 blocks were recorded. The order of the five velocities and the three stimuli was pseudo-randomized.

9.3.3 Results

A block was excluded from analysis if it was determined by visual inspection that more than half of the block of trials at each velocity for each image was unacceptable. This selection resulted in 160/165 or 97% valid block traces. The average velocity of each block (average of two blocks for the two subjects having 20 trials at each velocity) was

calculated. The total number of observations (smooth pursuit gains) then was equal to 130. The statistical analysis was done using MINITAB, with the data having been shown to satisfy the statistical assumptions of the ANOVA test (two-factor ANOVA with blocking on subjects).

To the end of testing the assumption of residual independence, Figure 75 is a plot of the residual values in the time sequence of the observations. As there was no regular pattern, the assumption of residual independence was taken to have been satisfied.

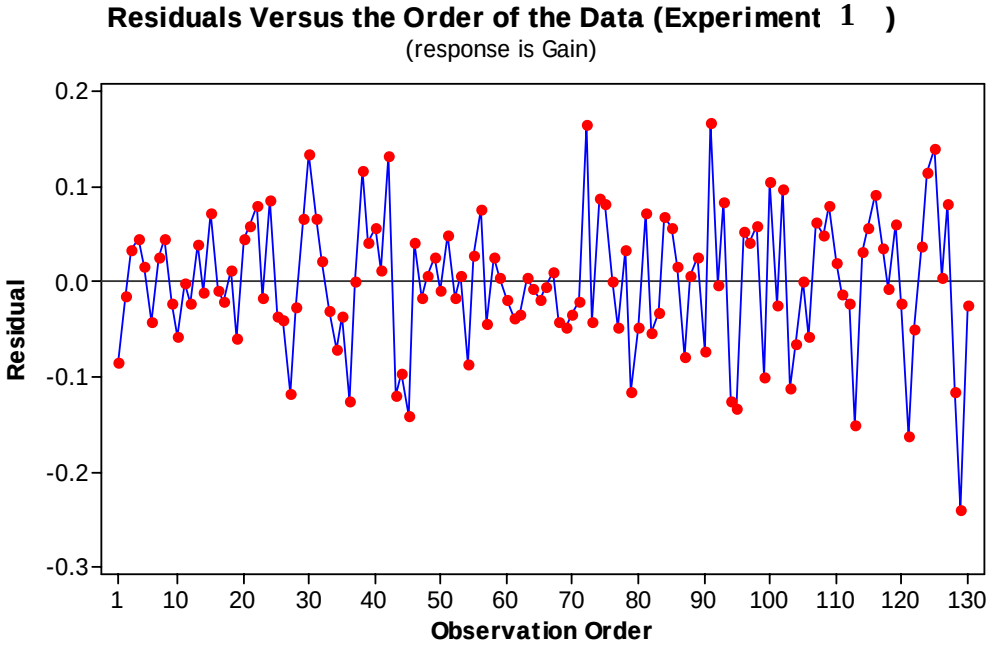

Figure 75. Check for residual independence (Experiment 1)

Also, it appears that the distribution of the residuals was normal based on Figure 76 (p>.05). As a result, the second assumption of normality in an ANOVA test was deemed satisfied.

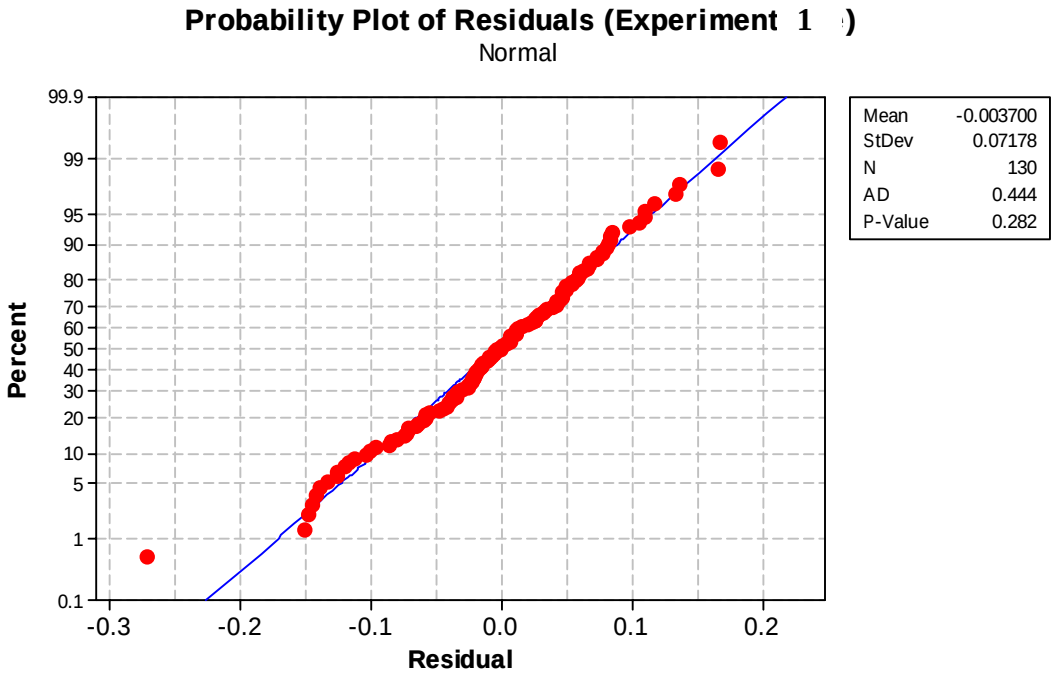

Figure 76. Check for normality (Experiment 1)

In the test of variance homogeneity, Bartlett's test and Levene's test showed conflicting results (Figure 77). Because there was considerable variation in the eye velocity data within and between individuals during measurements, this contradiction was overlooked and the ANOVA test was continued. Such smooth pursuit gain variation is common under similar conditions - for example, Meyer et al. (1985), van den Berg et al. (1986), and Heinen et al. (1998).

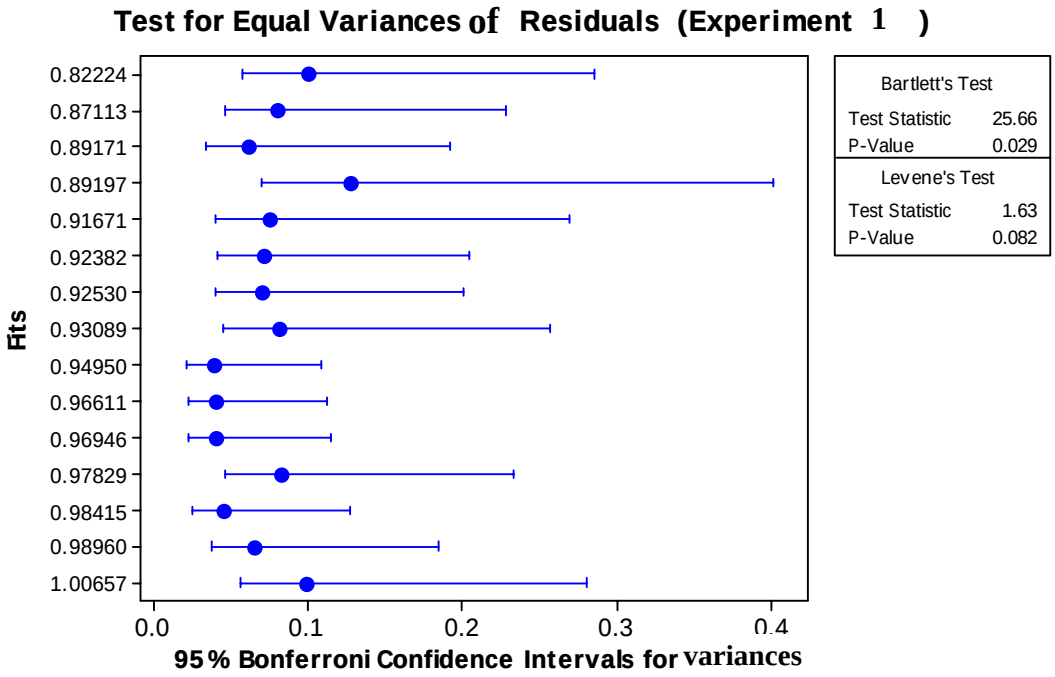

Figure 77. Check for equal variances (Experiment 1)

As shown in ANOVA Table 8, the stimulus and velocity both had statistically significant influences on the smooth pursuit gain ($p<.05$), however, there was no significant interaction ($p>.05$).

Table 8. ANOVA table (Experiment 1)

Factor	Degree of Freedom	SS	MS	F	P
Stimulus	2	0.038	0.019	4.81	0.01*
Velocity	4	0.23	0.058	14.51	0.00*
Subject	8	0.22	0.028	6.96	0.00*
Interaction (Stimulus×Velocity)	8	0.04	0.005	1.28	0.26
Error	107	0.42	0.004		
Total	129	0.96			

$p < .05$

Figure 78 shows the relationship between the smooth pursuit gain (averaged over nine subjects and three stimuli) and the target velocity. The gain decreased monotonically with the increase in the target velocity, and this relationship held for the target velocities ranging from 16 to 78 deg/sec. The increase in the target velocity also caused more velocity variation, as indicated by the increasing standard error.

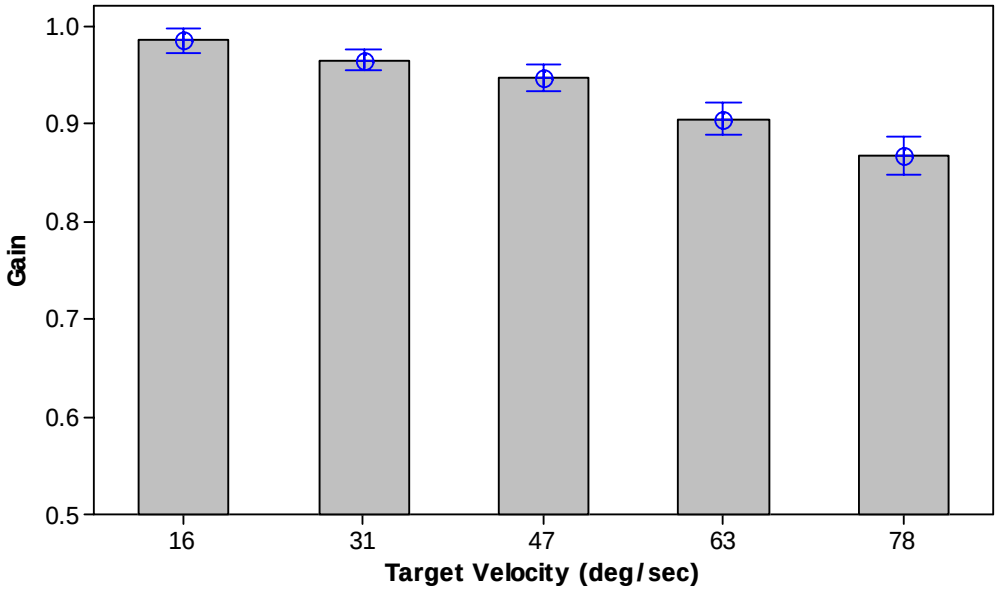

Figure 78. Gain versus target velocity (Experiment 1)

The gains for the three stimuli are shown in Figure 79. The gain for the dot image was significantly less than those for the apple and flower images. We used the shortest significant range (SSR) test to compare the individual levels among the three image stimuli. Since there was no statistically significant interaction between the velocity factor and stimulus factor, the images over all velocities were averaged together for comparison (Barker, 2005).

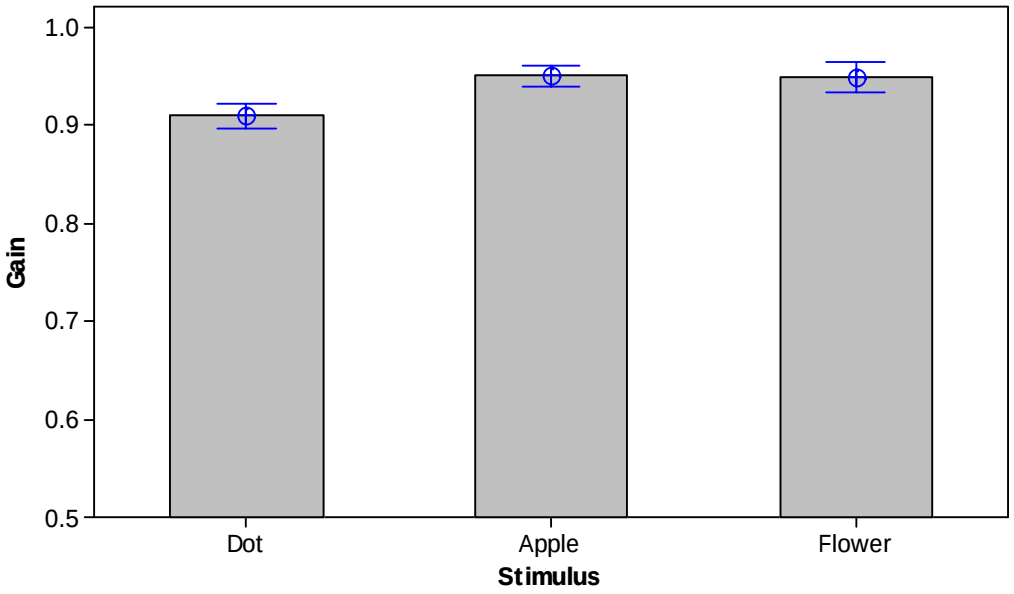

Figure 79. Gain versus stimulus (Experiment 1)

Comparison of the differences and the critical value in Table 9 show that the difference between the flower and dot images was statistically significant, while the difference between the flower and apple images was not.

Table 9. Shortest significant range (SSR) for comparing gain differences (Experiment 1)

Image	Mean of Gains	Difference between two levels	SSR (critical value)
Dot	0.910		
		0.038	
Flower	0.948		0.026
		0.001	
Apple	0.949		

The relationship between the eye velocity and target velocity is plotted in Figure 80. At the lowest target velocities (16 and 31 deg/sec), the eye velocities for the three images are grouped together. The velocity of the eye is greater for the apple image than for the flower or dot image at the intermediate target velocity of 47 deg/sec, while eye velocities for the dot were slower than those for the apple and flower at the two highest velocities (63 and 78 deg/sec).

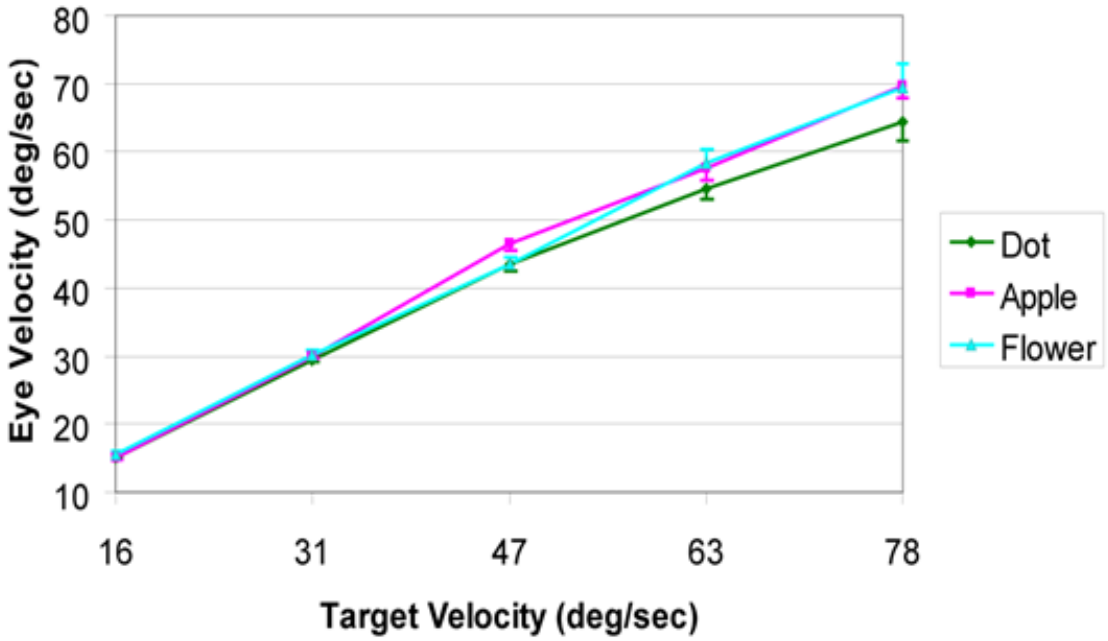

Figure 80. Eye velocity versus target velocity in Experiment 1. Error bars represented one standard error.

9.4 Experiment 2: Effects of Stimulus Size on Pursuit Velocity Limit

9.4.1 Subjects

Six subjects (students and faculty at the Rochester Institute of Technology) participated in this experiment. All subjects had normal or corrected-to-normal vision. Subjects provided informed consent and received an honorarium for participating. There were two females and four males from 19-43 years old (average 26.2). Two of the six subjects had also participated in Experiment 1.

9.4.2 Data Collection

In this experiment, the eye tracker was calibrated before the experiment and checked at regular intervals thereafter. No break was given to subjects. Three images were presented at two different velocities: 94 and 112 deg/sec. For each image and velocity, ten trials (one block) were conducted. As a result, data for a total of 360 eye-position traces in 36 blocks were collected. The order of two velocities and three stimuli were pseudo-randomized.

9.4.3 Results

The same criterion as in Experiment 1 for removing bad eye-trace blocks was used in this experiment, which resulted in 35/36 or 97.2% valid block traces. The average eye velocity in each block was calculated, and the corresponding gain value was used as the observation (total 35) in the following statistical analysis using MINITAB.

Statistical assumptions for applying the ANOVA test appear to have been satisfied. As shown in Figure 81, the assumption of residual independence was apparently satisfied since no regular pattern was observed in the plot. The assumption of normality in an ANOVA test also appears to have held ($p>.05$) based on Figure 82. Both Bartlett's test and Levene's test (Figure 83) showed an apparent homogeneity of variances ($p>.05$).

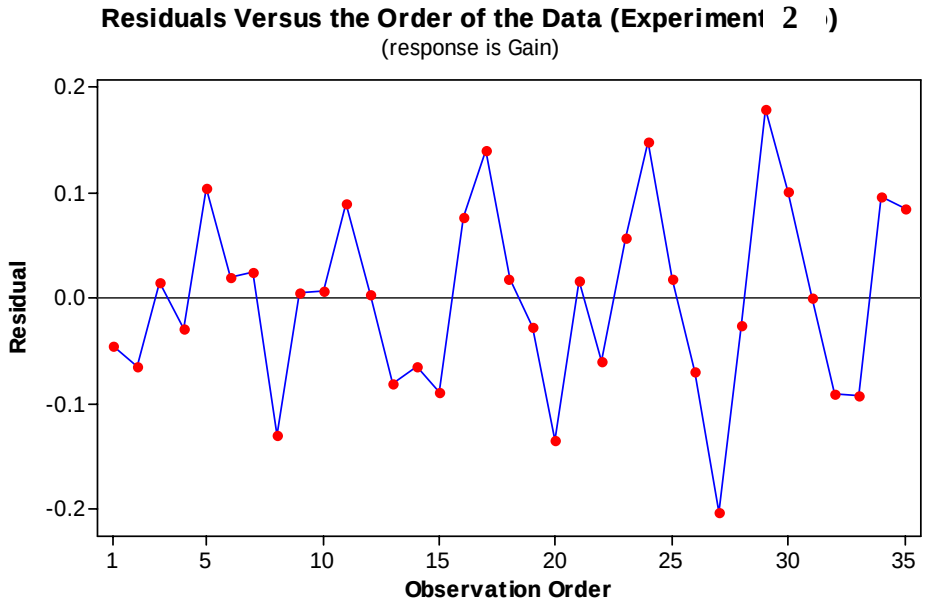

Figure 81. Check for residual independence (Experiment 2)

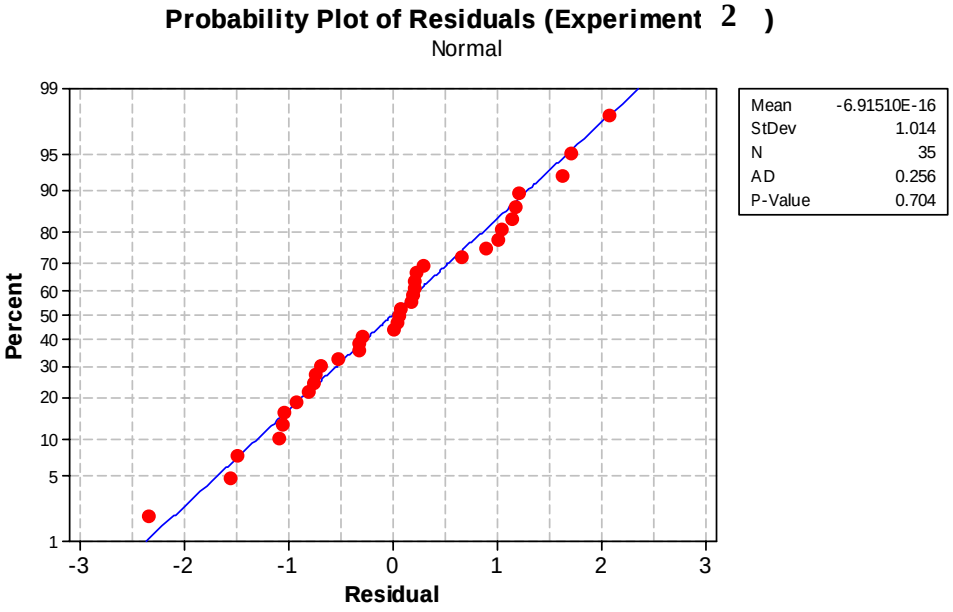

Figure 82. Check for normality (Experiment 2)

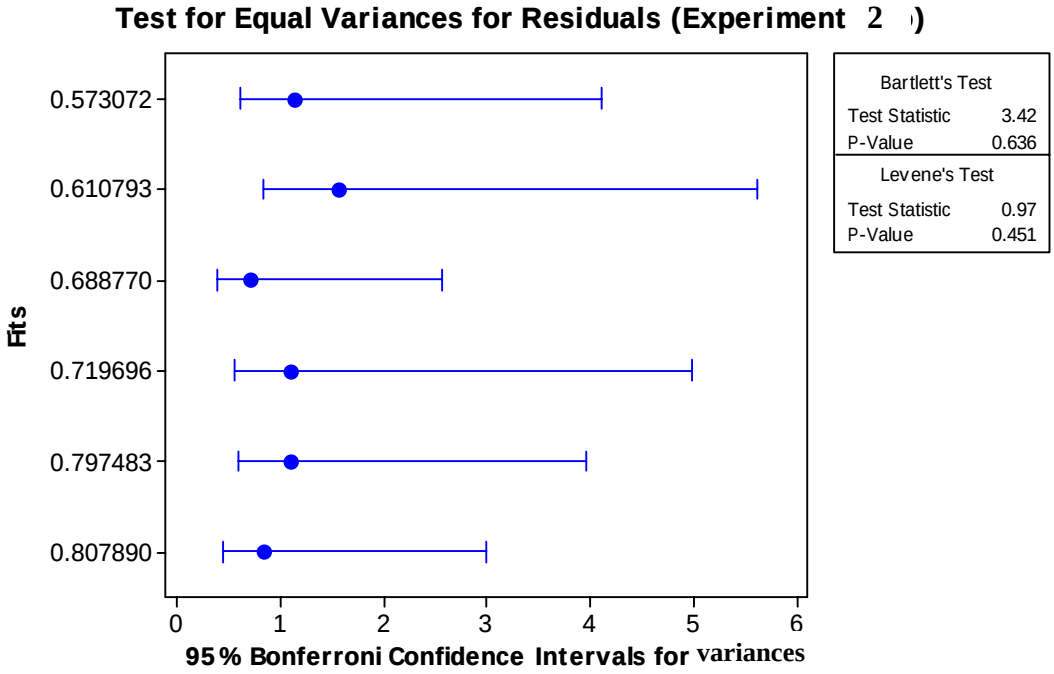

Figure 83. Check for equal variances (Experiment 2)

As shown in ANOVA Table 10, the stimulus and velocity both had statistically significant influences on the smooth pursuit gain ($p<.05$), while there was not a significant interaction of the two factors ($p>.05$). This result shows agreement with that from the Experiment 1, where relatively lower target velocities were used.

Table 10. ANOVA table (Experiment 2)

Factor	Degree of Freedom	SS	MS	F	P
Stimulus	2	0.08	0.04	10.76	0.00*
Velocity	1	0.16	0.16	41.58	0.00*
Subject	5	0.17	0.03	8.66	0.00*
Interaction (Stimulus×Velocity)	2	0.016	0.008	2.15	0.14
Error	24	0.092	0.004		
Total	34	0.53			

*$p < .05$

Figure 84 shows the smooth pursuit gains averaged over six subjects and three stimuli as a function of target velocity. The gain decreased dramatically when the target velocity increased from 94 deg/sec to 112 deg/sec (gain dropped from 0.76 to 0.63), while the standard error increased slightly (0.02 to 0.03).

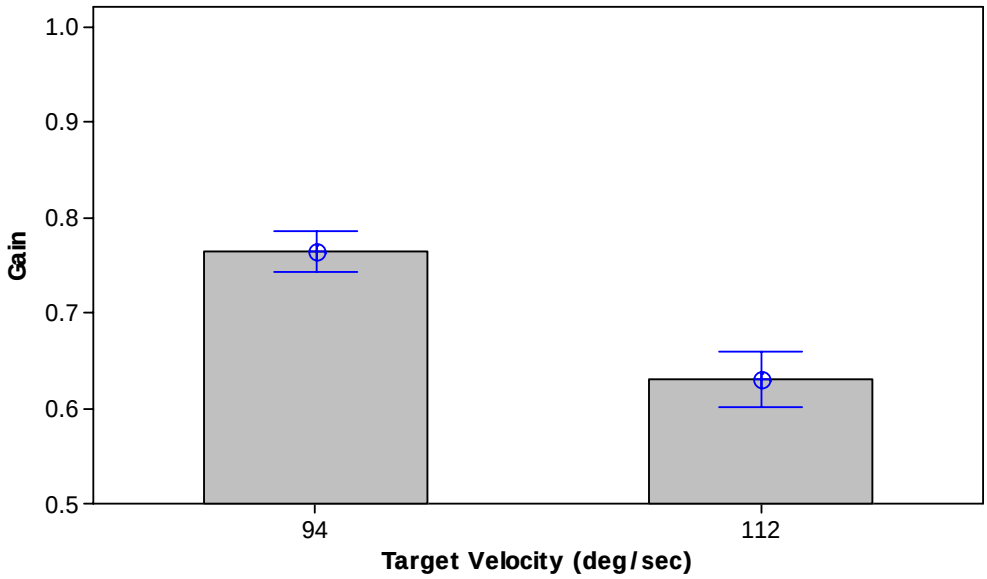

Figure 84: Gain versus target velocity (Experiment 2)

As before, the individual levels among the three image stimuli were analyzed by calculating the SSR. The results are shown in Table 11. Again, the difference between the dot image and the other two images was statistically significant while the difference between the flower and apple image was not.

Table 11. Shortest significant range (SSR) for comparing gain differences among images (Experiment 2)

Image	Mean of Gains	Difference between two levels	SSR (critical value)
Dot	0.631	0.078	
Apple	0.709		0.0533
		0.053	
Flower	0.762		

In Figure 85, it is evident that the average gain of the three stimuli decreased with the increase of the stimulus extent.

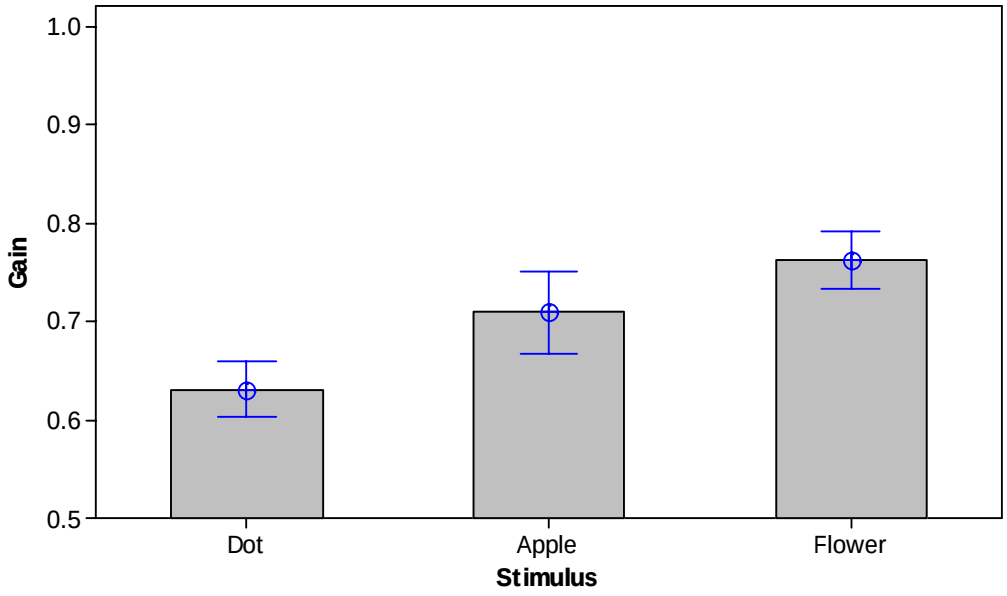

Figure 85. Gain versus stimulus (Experiment 2)

To study the upper velocity limit of smooth pursuit, data sets from Experiments 1 and 2 were combined. Note that the standard error from these two experiments cannot be directly compared because of the different number of subjects. The relationship between the gain and target velocity after pooling the two data sets is plotted in Figure 86. Overall, the smooth pursuit gain decreased with the target velocity. The gain became considerably lower (about 0.6 for the dot and apple and 0.7 for the flower) at the highest target velocity. Additionally, the gain decrements for the apple and flower were lower than that for the dot.

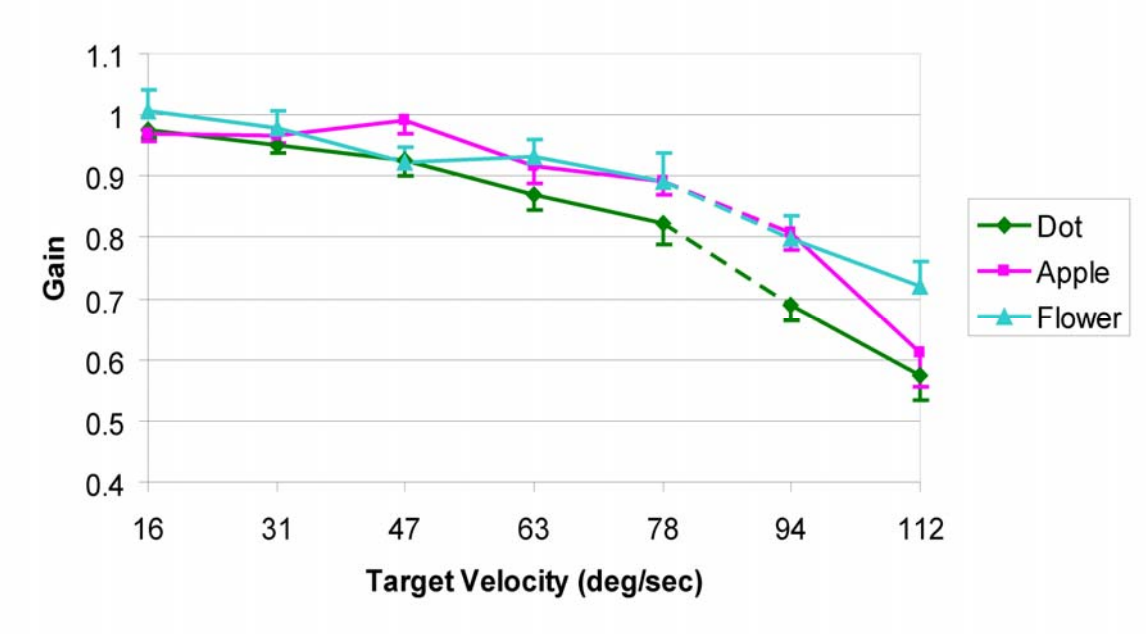

Figure 86. Gain versus target velocity after pooling the data sets in two experiments. The dashed lines separate the two data sets. Error bars represented one standard error.

The relationship between the eye velocity and target velocity after pooling the two data sets is plotted in Figure 87. The saturation of the eye velocity for the two extended stimuli was slower than that for the small dot; the velocity saturation of the dot appeared

at the target velocity of 78 deg/sec, while that of the apple showed up at 94 deg/sec. The eye velocity of the flower kept increasing even at the highest target velocity. Furthermore, the eye velocity of the apple at the target velocity of 94 deg/sec was comparable to that of the flower, while the eye velocity decreased to a level close to the dot when the target velocity increased to 112 deg/sec.

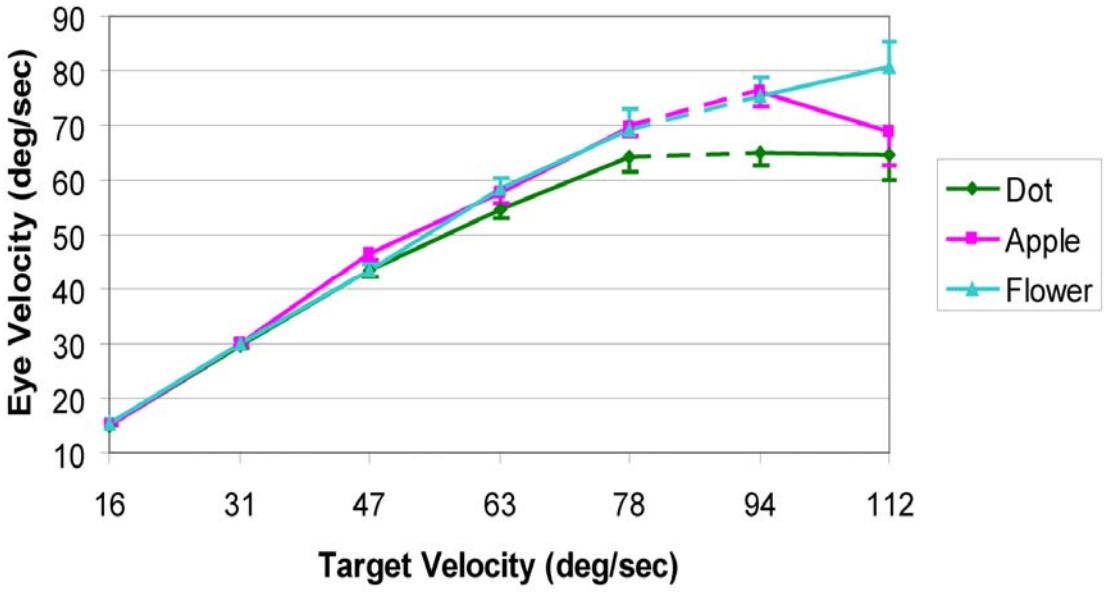

Figure 87. Eye velocity versus target velocity after pooling the data sets in two experiments. The dashed lines separate the two data sets. Error bars represented one standard error.

9.5 Discussion

9.5.1 Effect of Target Velocity

The smooth pursuit gain decreased with the increase in stimulus velocity, as has been shown in other studies using synthetic stimuli (Collewijn, Steinman, et al., 1985; van den

Berg & Collewijn, 1986). This effect may be interpreted by the general mechanism of the smooth pursuit system. In the pursuit visual pathway, the velocity signal of the target is integrated neurally into an eye position signal in the oculomotor brainstem, and the combination of these two signals is sent to the extraocular muscles (Pola & Wyatt, 1991). The higher the target velocity is, the shorter the integration time available, which may cause more inaccuracies in generating control signals for pursuit eye movements.

9.5.2 Effect of Stimulus Size

There are a number of potential interpretations as to why a large stimulus benefits the smooth pursuit performance.

(1) An extended target may activate more velocity detectors (Heinen & Watamaniuk, 1998; van den Berg & Collewijn, 1986). As stated in Section 9.1, during a smooth pursuit task, a retinal slip signal is sent as a feedback to the oculomotor system to control eye movements. When tracking an extended target, more feedback resulting from the retinal slips will be generated, and thus more velocity sensors in the oculormotor system would be activated simultaneously. On the other hand, as shown in some other studies (Masson, Proteau, et al., 1995; Niemann & Hoffmann, 1997), smooth pursuit velocities were significantly suppressed in the presence of a stationary textured background. These results imply the activated velocity sensors may provide not only central excitatory signals, but also surround inhibitory signals (Heinen & Keller, 2004). This center-surround mechanism is well known in some visual cells in the retina and visual cortex.

(2) Motion signals from inside and outside the central fovea are summed or averaged.

Because the extended stimuli (subtending 2° and 17° horizontally) used in the present study were projected onto an area of the retina that is larger than the size of the isoacuity area in the fovea (about 1°), it is possible that the motion signals activated in the parafovea and peripheral retina are added to (or weighted-averaged with) those triggered in the central fovea. Therefore, the information collected from the central fovea is not necessarily providing sole cues to direct the smooth pursuit eye movement; those from other parts of the retina may also play some role. Furthermore, a stronger sensory stimulus may provide a larger integration (velocity) signal pooled from all activated components and, hence, lead to a better estimation of the target velocity. Increasing the angular extension of the stimulus from the dot to the apple image, which was large enough to cover the fovea, caused a significant increase in smooth pursuit gain. Increasing the image size beyond the fovea (from the apple to the flower image), however, did not lead to further significant increases in gain. Further work with stimuli of different angular extent is needed to investigate this hypothesis.

(3) There could be a confounding effect with the OKR. Optokinetic response (OKR) is an eye movement similar to smooth pursuit, except that it is usually considered as induced by a large- or full-field stimulus (Collewijn, 1991). Some researchers have even suggested that the distinction between smooth pursuit and OKR is not clear (Heinen & Keller, 2004). Animals without foveae, like rabbits, use OKR to stabilize retinal motion, while animals with foveae rely on smooth pursuit extensively. The same effect of extended stimuli on the OKR has also been found in the studies by using full-field stripe patterns (Collewijn, 1991). Heinen and Watamaniuk (1998) proposed that the evolution of the optokinetic system produced the smooth pursuit system, which extended the ability of voluntary pursuits in foveate creatures. The smooth pursuit system may simply be a

subsystem of the optokinetic system (or the reverse), since they excite basically the same areas in the brain (Carpenter, 1988). Activating more visual sensors in the whole system generates a more accurate signal guiding the pursuit eye movement.

9.5.3 Limit of Smooth Pursuit Velocity

The smooth pursuit gains of the dot image were 0.82 at the target velocity of 78 deg/sec, 0.69 at 94 deg/sec and 0.57 at 114 deg/sec. This results in a saturated eye velocity of 64 deg/sec for the dot image, which is within the range reported in Collewijn, Steinman, et al. (1985) and lower than that in Meyer, Lasker, et al. (1985). Both of these two studies used a small target. The discrepancy may be due to different experimental setups and individual variations found in both the above two studies and the present one. Moreover, the limit of smooth pursuit velocity was found to be associated with the stimulus size in our study; an extended stimulus could boost the smooth pursuit velocity limit. In Figure 87, the eye velocity for the apple was close to that for the flower at the target velocity of 94 deg/sec, while it dropped to a level close to the dot at the target velocity of 112 deg/sec. In other words, the apple image was tracked like the full-field flower image at 94 deg/sec, while at 112 deg/sec, it degraded to the same performance as the dot image. This suggests that the beneficial effect of an extended stimulus may become limited at extreme velocities. Also, note in the present study and in Meyer, Lasker, et al. (1985) (not clear in Collewijn, Steinman, et al., 1985), the subject's head was stabilized by a chinrest, whereas people may move their head to supplement pursuit eye movements in real viewing conditions. These supplementary movements are especially crucial when tracking quick-phase, large-range motions. This suggests that free head movements to

track may result in a considerably higher limit of pursuit velocity than those generally observed in the present and previous studies.

Chapter 10 Eye Tracking Applications: Smooth Pursuit Dynamics in Unconstrained Vibrating Conditions

In this chapter, the video-based eye tracking technique was applied in the second eye movement study, which investigated the gaze behavior of passengers tracking an object while riding in a car. In contrast with the experiments in Chapter 9, the head movements of the subject were unconstrained and the subject was exposed in a whole body vibration condition when performing the visual task. Furthermore, the visual stimulus was presented close to the eye, while it was shown at a distance in the previous chapter. This chapter concludes with a discussion of different smooth-pursuit dynamics between the controlled static condition and in a moving car with free head movements.

10.1 Background

10.1.1 Whole Body Vibration in Vehicles

Commuters read newspapers or use mobile devices when taking buses or trains.

Passengers reading or watching TVs are commonly observed on airplanes, which can be very bumpy. In automobiles, children often sit in back seats playing video games or watching videos when traveling with their parents. In these transportation vehicles, people are often exposed to the motion called *whole body vibration* (WBV), occurring when the mechanical vibration from the vehicle is transmitted to the occupants. The effects of the WBV depends on a number of factors, such as the characteristics of the motion (direction, frequency, magnitude, etc.), the transmissivity of supportive media (e.g., seat and floor), and physiological (e.g., mass and posture) and psychological status of the occupant. The whole body vibration is different from a local vibration in the way that its effect is not local to any particular points of contact.

There are six axes for vibration as shown in Figure 88, or the six degrees of freedom in motion: fore-to-aft (x-axis), lateral (y-axis), vertical (z-axis) and rotational movements about the x-, y- and z-axes (called roll, pitch and yaw, respectively). The rotational movements around x-y-z axes are considered irrelevant to the prevalent form of vibration in vehicles (Griffin & Lewis, 1978) and therefore are infrequently discussed in the literature of WBV. The magnitudes of lateral (fore-aft and left-right) vibrations are considerably greater than that of vertical vibrations for producing the same reading errors or reading times in visual tasks (Lewis & Griffin, 1980). However, the vertical vibration has the most adverse effects on visual performance because the resonance of the head in the vertical axis is the most prominent. When the resonance happens, at the frequencies between 2 and 10 Hz (Griffin, 1992), the vibration is transmitted from the head to the eye and thus causes degradation of visual performance. The resonance of the eye within the head happens at the high frequencies between 20 and 70 Hz (Griffin, 1975; Griffin, 1981; Stott, 1984), although it is largely variable within the same individual on different

occasions (due to body posture, seating characteristics, etc.) and between individuals on the same occasion. The fore-aft vibration can be more detrimental to vision by adding a backrest, which increases the transmission of the vibration from the vehicle to the head/eye (Lewis & Griffin, 1980).

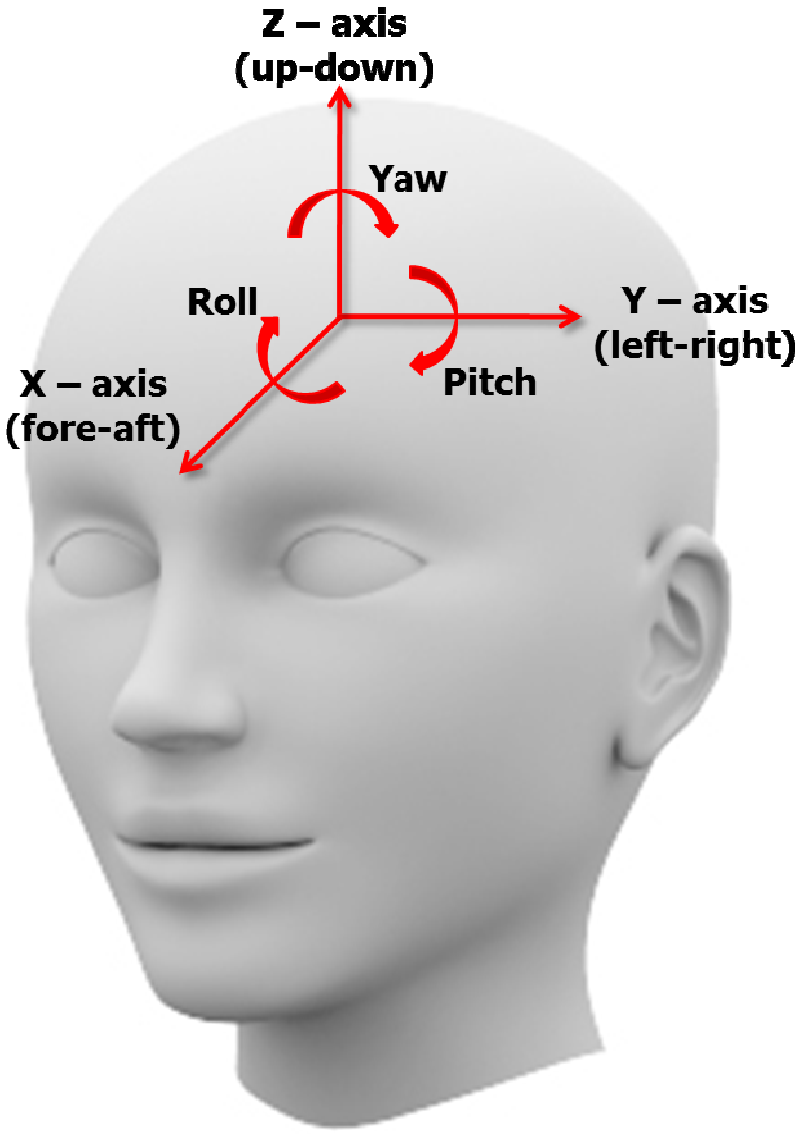

Figure 88. Six degrees of freedom in WBV

The Root Mean Square (RMS) value of acceleration, in units of m/s^2, is a widely used quantity for measuring vibration. The magnitude of vibration differs in different environments: for a railroad locomotive, the mean vibration levels are:: the fore-aft axis $x = 0.18\ m/s^2$, the left-right axis $y = 0.28\ m/s^2$, and the up-down axis $z = 0.32\ m/s^2$ (Johanning, Wilder et al., 1991); for an urban taxi, the mean is $0.31\ m/s^2$ for z-axis (Chen, Chang et al., 2003); for a highway transport truck on rough roads (parenthesized values are for smooth roads), the means for the x-y axes are equal at $0.24\ m/s^2$ ($0.12\ m/s^2$ for x-axis and $0.15\ m/s^2$ for y-axis), and the value for the z-axis is $0.52\ m/s^2$ $(0.35\ m/s^2)$ (Cann, Salmoni et al., 2004).

10.1.2 Vestibular System

The vestibular system is an essential component in the process of motor responses, which are crucial to mediate relevant movements of the human body. The outer layer of the system, called the bony labyrinth, is located in the inner ear and protects the underlying vestibular sensory structures. The membranous vestibular labyrinth, situated inside the bony labyrinth, consists of three semicircular canals and two otolith organs. The semicircular canals (lateral, superior and posterior) are responsive to rotational head movements (angular accelerations), and their orientations define the three coordinates in which the body senses angular rotations. The otolith organs signal translational head movements (linear accelerations, e.g., during running or turbulence in vehicles) or the tilting of the head relative to gravity (pitch and roll movements). These five vestibular receptor organs complement each other in function to transduce orienting and positional

stimuli into neural signals that are then sent to the brain.

The sensory receptors in the vestibular system are called hair cells, which are located in the crista of semicircular canals and in the macula of the otolith, respectively**.** During a head motion, the bony labyrinth, membranous labyrinth, and attached bodies of the hair cells turn with the head. However, the hairs are bent against the endolymph (the fluid that fills the space inside the membranous labyrinth), which tends to remain stationary due to inertia and friction with ducts (Pansky, 1988, page 300). When the motion stops or reaches a constant speed, the inertia and friction are overcome such that the hairs slowly return to their rest positions and the stimulation of the hair cells ceases.

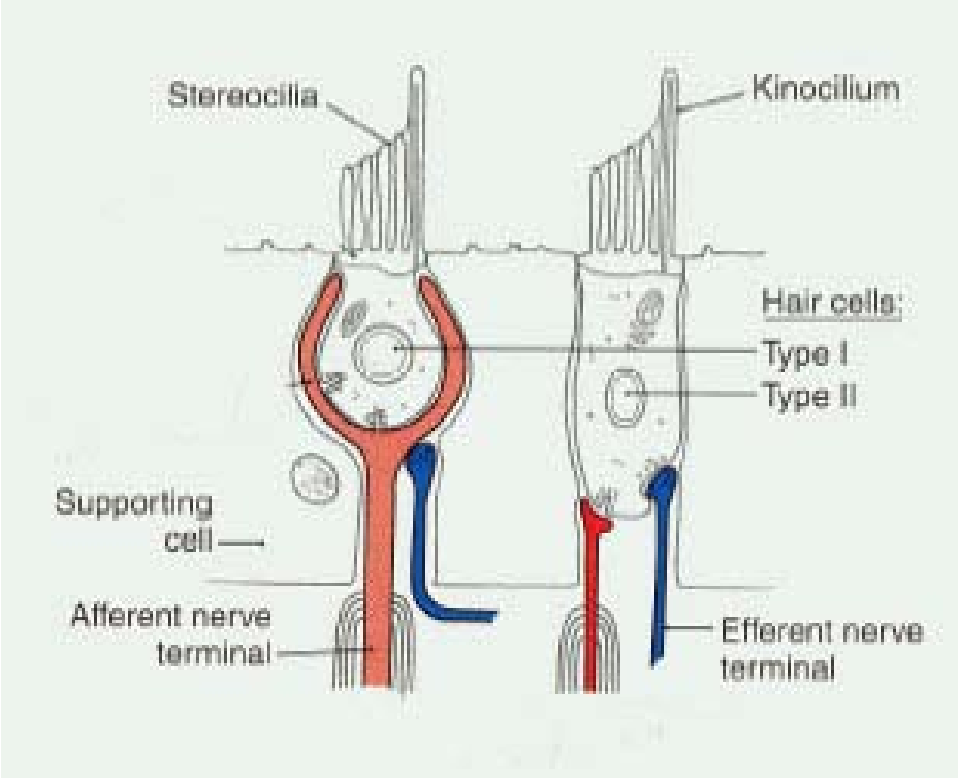

Figure 89. Vestibular sensory receptors. There are two types of hair cells and supporting cells. Each cell contains around 60-100 stereocilia (arranged in an ascending height) and one kinocilium. From Dickman (2006).

10.1.3 Visual Research in Body and Display Vibrations

When exposed to WBV in vehicles, people generally are observed to have degraded visual performance; tremendous retinal slip (the mismatch between the target and projected image on the retinal mosaic) happens in unstable conditions. It is also considered to be the error of compensatory eye movements; a perfect compensatory eye movement stabilizes the moving object on the fovea and therefore produces no retinal slip. The image mismatch in the retina may be caused by a number of factors, such as a small field of view (if using mobile devices), distorted images in the retina, and motion sickness (drowsiness, headache, and nausea).

Although there have been many studies on the legibility of vibrating displays, very few studies have addressed the case where the observer or both the observer and the display are vibrated (Griffin, 1992). Given the fact that vestibular eye movements are not engaged when only the display is in motion, the detrimental effects on visual performance is over-estimated when the display alone is in motion in comparison to the more natural case when head motion is also involved, because VOR eye movements are triggered to compensate for head movements. In one of the few studies comparing the effects of display vibration, whole-body vibration and simultaneous whole-body-and-display vibration on visual performance, display vibration alone had the largest performance decrement, followed by the body vibration alone and then the simultaneous whole-body-and-display vibration (Moseley & Griffin, 1986).

In the present eye tracking study, the observers' gaze patterns were recorded when they sat in a car, where both the body and viewing target were vibrated naturally by the

motion of the car. An attempt was made to answer the question as to whether the visual performance is different when people perform pursuit tasks in an unconstrained vibrating condition in comparison with in a controlled static laboratory. The hope is that these results will provide useful information in guiding engineers to build countermeasures to alleviate visual performance degradation in unstable environments.

10.2 Experiment 3: Effect of Whole Body Vibration on Smooth Pursuits in an Unconstrained Condition

10.2.1 Experimental Setup

Subjects were seated on the back seat of a middle-size sedan driving along a local (light traffic) route. The viewing distance was approximately 25 inches (63.5cm). The same dot and apple images (Figure 73), extending 0.1 and 2 degrees of visual angles, used in the studies of Chapter 9 were used in this study. The stimuli moved horizontally from the left to right on a 3.5" LCD mobile display (iPod Touch Generation 3), with a pixel pitch of 0.16 mm.

10.2.2 Eye-Movement Recording

Eye movements were recorded by a portable video-based RIT eye tracker (Section 5.5.2), which tracked monocular eye positions at the standard NTSC frame rate of 30 Hz (33.3 ms). Viewing was binocular. For each subject, before and after the experiment, the

calibration of the eye tracker was checked and recalibrated if necessary. The accuracy of the eye tracker within the calibration region is approximately 1º.

10.2.3 Subjects

Fourteen subjects (students and faculty at the Rochester Institute of Technology) participated in the experiment. All subjects had normal or corrected-to-normal vision. The subjects provided informed consent and received an honorarium for participation. Three subjects had unacceptable amounts of tracking loss and one subject felt nauseous during the experiment, so those data were excluded. Consequently, eye-tracking data from ten subjects were used in the statistical analyses.

10.2.4 Experimental Procedure

The subjects were asked to track the visual stimuli (dot and apple, defined as two sessions) as smoothly as possible. For the apple image, they were instructed to pick any point on the image to track, but to be consistent within each tracking trial.

Five target velocities were used in this experiment, namely, 2, 4, 8, 12 and 16 deg/sec. Due to the limitation of the frame rate (30 Hz) of the mobile display, a high target velocity (> 16 deg/sec) would have caused double-images similar to those shown in Figure 42. Therefore, a target velocity higher than 16 deg/sec was not considered in the experiment. At each target velocity in each session, data for 10 trials (one block) were collected. As a result, a total of 1000 eye-position traces in 100 blocks from 10 subjects were analyzed. The order of the five velocities and two stimuli was pseudo-randomized.

10.3 Eye Tracking Data Analysis

The eye gaze positions in the scene image were output from the eye tracker software (Yarbus 1.0), while the target (apple) positions were coded manually frame by frame from the scene-and-eye-combined videos (Figure 90). Due to the small size of the dot image and its low contrast with the background, the motion of the dot could not be detected from the videos (i.e., signals are lost). Therefore, only the data from the apple session were studied in the following section, and the results are discussed at the end of the chapter.

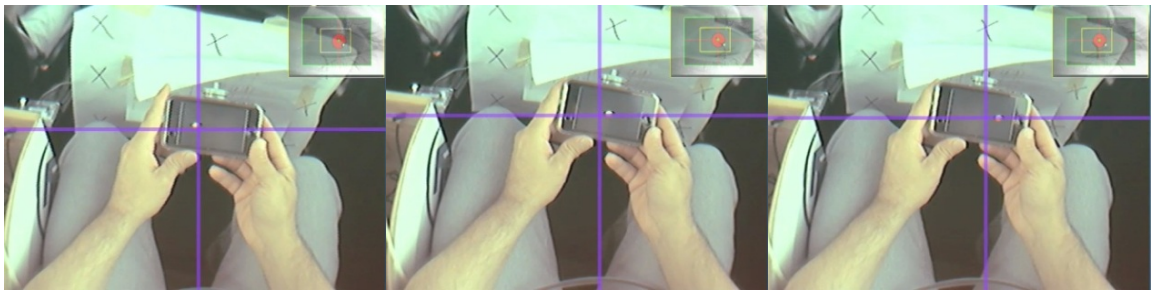

Figure 90. A subject tracking the apple moving at the velocity of 4 deg/sec. (a) the target running on the left of the screen; (2) after 0.83 sec (25 frames); (c) after 1.67 sec (50 frames). The cross represents the eye fixation point in the scene.

Like the smooth-pursuit study in the controlled laboratory condition in Chapter 9, eye-position data from each trial was plotted and analyzed visually. An example eye trace in the study of tracking the apple is shown in Figure 91. A trial was excluded from analysis if there were significant track losses in the trace. This resulted in 419/500 or 83.6% valid traces.

The calculations of target and eye velocities were done using a similar approach as in Chapter 9. If a value exceeded three standard errors from the mean, it was considered an outlier and excluded from the statistical analyses. This additional criterion led to 228 valid trials in the end (the data appearing normal distributed as shown in Figure 93) or 45.6% of the total traces initially recorded. The distribution of valid trials by target velocity and by subject is listed in Table 12. The low rate of the valid traces was due to: (1) tracking losses due to vibration in the car, and (2) measurement errors in determining appropriate segments in calculating target and eye velocities, especially at high target velocities (having limited samples). The target or eye velocity in each block was averaged, which resulted in 50 observations (i.e., smooth pursuit gains) in total for 10 subjects at 5 target velocities.

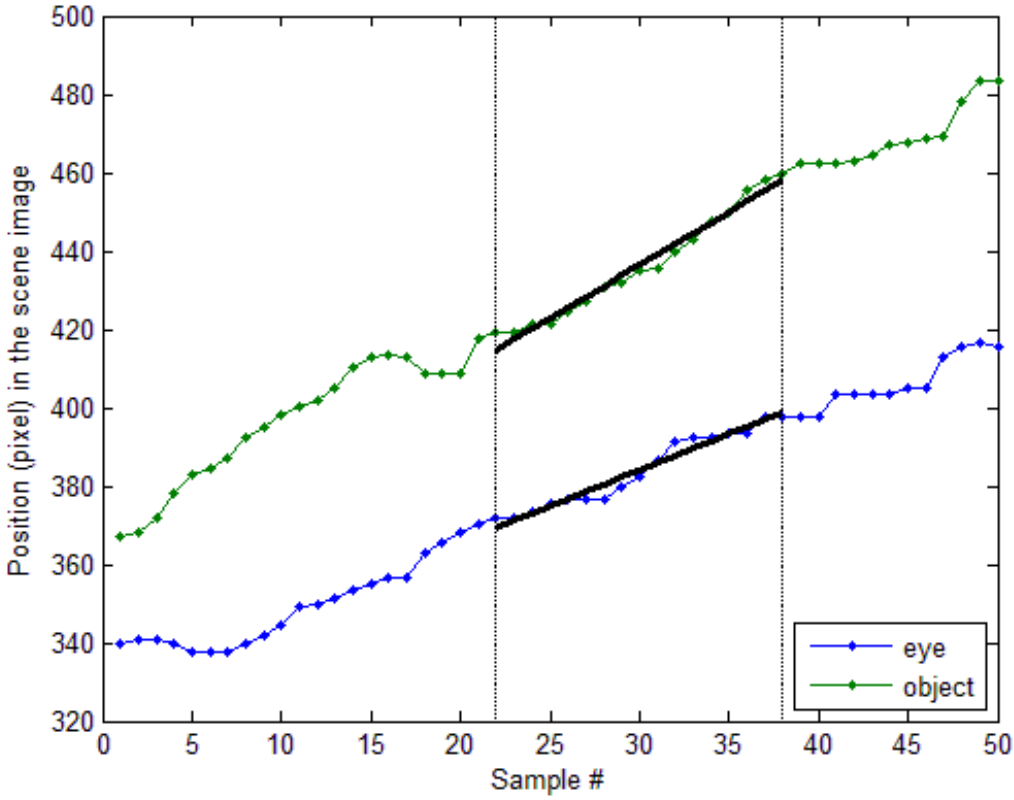

Figure 91. An example eye-position trace in Experiment 3. Two dashed vertical lines depicted a stable pursuit segment within which eye-position samples were linearly fitted and a slope (solid, slanted segment) was calculated as the velocity for the smooth pursuit.

Table 12. The number of valid trials in the eye-tracking-in-car study

target velocity	valid trial #
2	36
4	35
8	44
12	55
16	58
total	228

subject	valid trial #
1	39
2	15
3	23
4	18
5	23
6	28
7	25
8	17
9	16
10	24
total	228

10.4 Results

The data were determined to satisfy the statistical requirements of the ANOVA test. As shown in Figure 92, the assumption of residual independence appears to have been satisfied since no clear pattern was observed in the plot. The assumption of normality in an ANOVA test also appears to have held ($p>.05$) based on Figure 93. Both the Bartlett's test and Levene's test (Figure 94) showed the homogeneity of variances ($p>.05$).

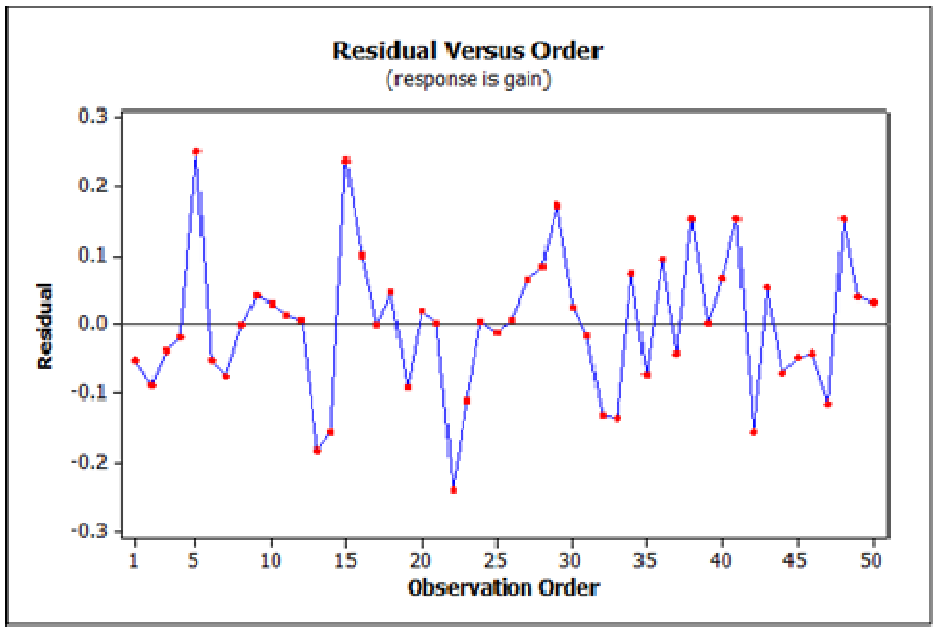

Figure 92. Residual vs. order plot in the study of pursuit eye movements in car

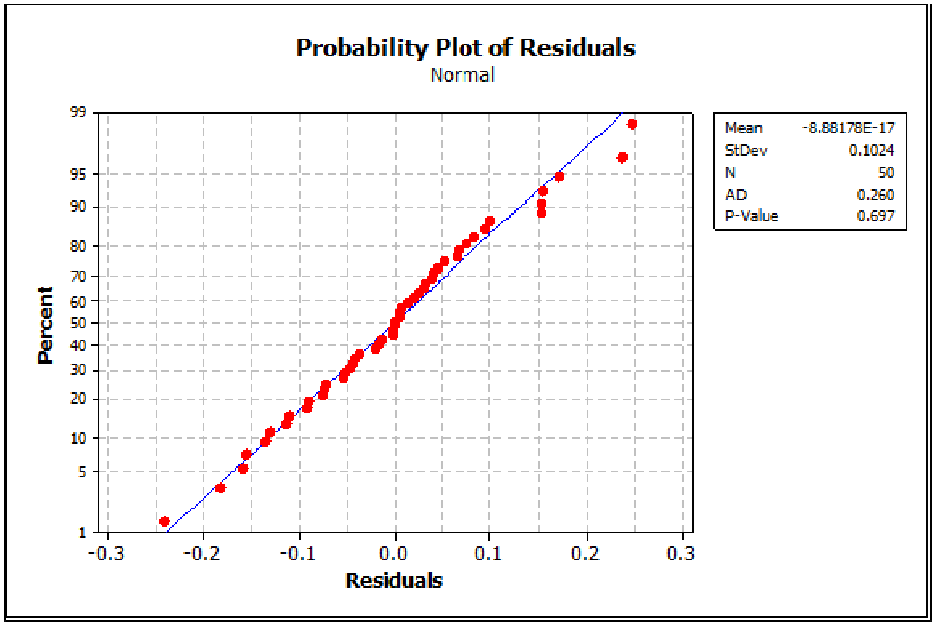

Figure 93. Probability plot of residuals in the study of pursuit eye movements in car

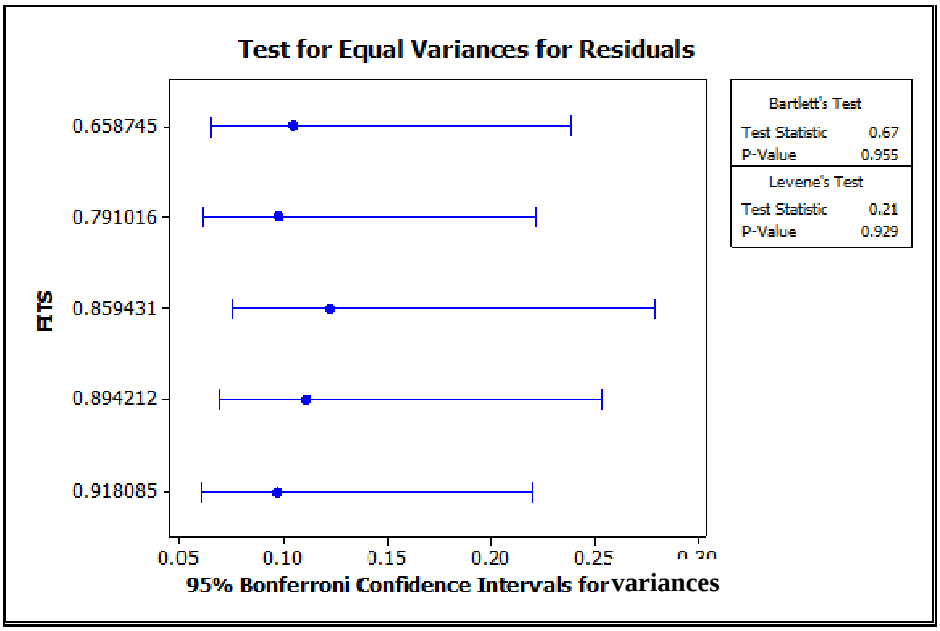

Figure 94. Plot of test for equal variances in the study of pursuit eye movements in car

The difference of smooth pursuit gains for 5 target velocities were statistically significant ($F_{4,45} = 9.5$, $p<0.001$) in a one-way ANOVA (Table 13). The smooth pursuit gains averaged over 10 subjects as a function of target velocity were shown in Figure 95. Overall, the smooth pursuit gain decreased with the increase of target velocity. The gain was less than 0.7 - unity means perfect tracking - at 16 deg/sec, and significant retinal slips happened when the subject tracked a rapidly moving target in a car.

Table 13. ANOVA table (experiment 3)

Factor	Degree of Freedom	SS	MS	F	P
Velocity	4	0.43	0.11	9.50	0.00*
Error	45	0.51	0.011		
Total	49	0.95			

$p < .05$

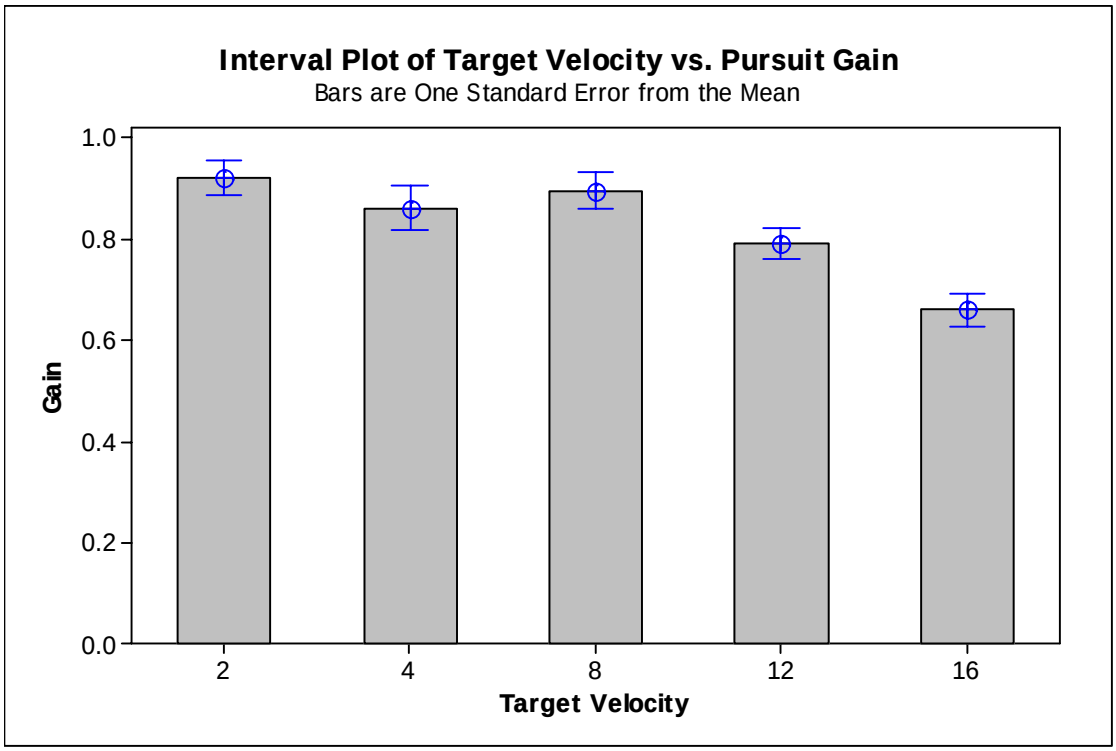

Figure 95. Target velocity vs. smooth-pursuit gain for the eye tracking study in car

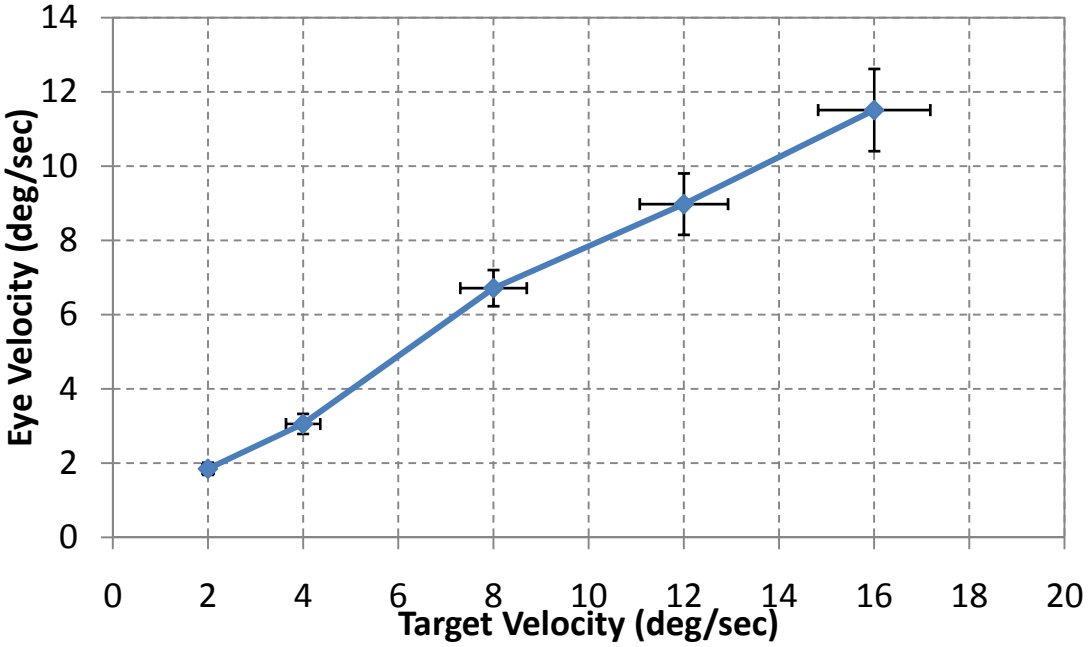

Figure 96. Target velocity vs. eye velocity in the experiment of pursuit eye movements in car. Error bar represents one standard error from the mean. Error bar in the eye velocity (y-axis) is due to variation within and between the subjects. Error bar in the target velocity (x-axis) is due to measurement error, caused by head movements and vibration in the car.

A multiple comparison test ('multcompare' function in MATLAB Statistical Toolbox) was used to check which pair of target velocities had a significant difference in gains. As shown in Figure 97, only the smooth pursuit gains at 2, 4, and 8 deg/sec were statistically different from that at 16 deg/sec.

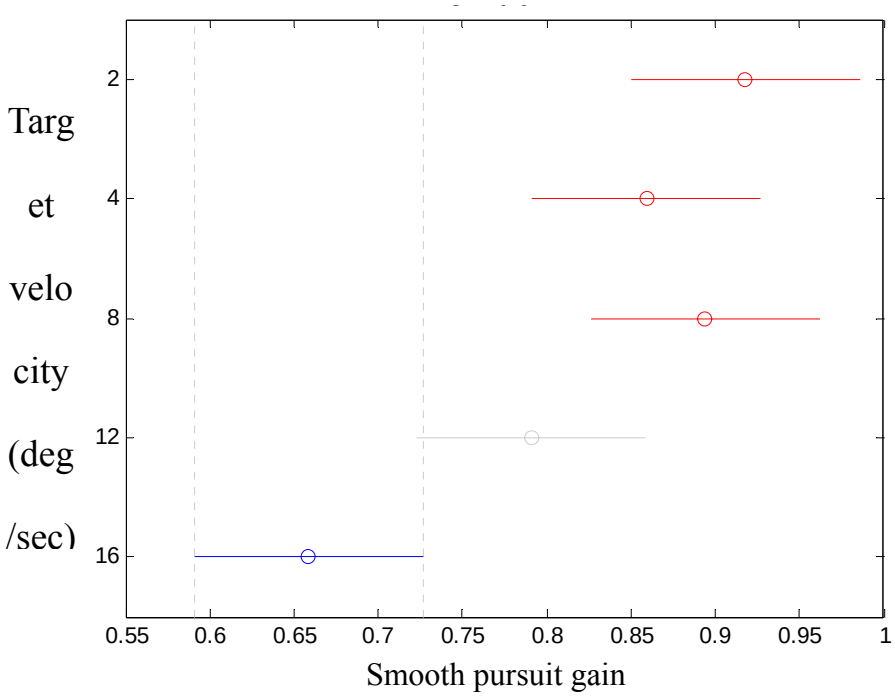

Figure 97. Smooth pursuit gain with comparison intervals. Circles represent mean gain values and horizontal lines represent 95% confidence intervals. Pairs with significant difference in gains are highlighted in the plot (blue line vs. other lines).

10.5 Discussion

In general, the smooth pursuit gains decreased with an increase in the target velocity in the study of smooth pursuit eye movements in a car. The smooth tracking became harder when the target moved faster. The gains from this study ranged from 0.67 ~ 0.9. At a very low target velocity, such as 2 deg/sec, the gain value was still less than unity; the pursuit was not very smooth in the vibration condition even when the subjects tracked a slowly moving target. On the other hand, the gain value at the highest target velocity, namely, 16 deg/sec, was 30% and smaller (gain < 0.7) in tracking smoothness than that in a perfect

tracking scenario (gain = 1); the mismatch between the moving target and its projected image on the fovea was clearly evident.

Between the two extreme target velocities (lowest and highest), although the smooth-pursuit gains decreased with the increase of target velocity, the difference was insignificant (as shown in Figure 97). The challenge of tracking an object smoothly in a moving vehicle and the possible contamination of translational eye movements (eye tracker headband slippage, etc.) caused high variations in the measured eye velocities. Furthermore, as shown in the horizontal error bars in Figure 96, the measured target velocities also involved uncertainty. The variation was marginal at low speed motions, while it became significant at high speeds due to the display motion caused by the WBV and the limited frames available to determine the target velocities.

Compared to the smooth-pursuit gain measured in the controlled static condition in Chapter 9, the value at the same target velocity in the unconstrained vibrating condition in this chapter was much lower. The gain at 16 sec/deg was 0.67 in the present study while it was 0.98 in Experiment 1 (Section 9.3). That means the negative impact of vibrating on visual performance was considerable, even though free head motions and compensatory eye movements (VOR) by the vestibular system were supposed to counteract the degradation to some degree. The resonances of the head and the eye within the head in the whole-body vibration condition greatly impair the smoothness of the eye in executing a pursuit task.

Chapter 11 Conclusions and Future Work

An object in a scene is perceived in great detail because we move the eye to focus it on the fovea, a region with highest acuity in the retina. As the eye is directed to the region in the scene that is interesting and/or important to the observer, tracking the eye movements provides a way to study what draws the person's interest and how visual information is collected. Video-based eye tracking is a popular technique used to track eye movements. This dissertation is primarily focused on optimizing the video-based eye tracking techniques and applying eye-tracking systems to study the dynamics of smooth-pursuit eye movements.

11.1 Eye Tracking Technique Optimizations

Based on how an "eye camera" is positioned, the video-based eye tracking technique can be categorized into two types: remote and head-mounted. Remote trackers have the eye camera placed (commonly on a desk) at a distance from the observer, whose gaze behavior is recorded during a visual task. Head-mounted video-based eye trackers have the camera attached to headgear which is worn by the subject. Head-mounted eye

tracking systems allow subjects to perform natural tasks under minimum restraint by the apparatus. With the improved flexibility of head-mounted trackers, specifically those mounted on eyeglass frames with light weight (e.g., RIT Light Weight Eye Tracker), comes the possibility that eye tracker movements will undesirably affect the tracked eye-in-head (angular orientation) output. This unavoidable effect is the result of tracked eye features (the pupil and CR) being displaced in the eye images due to translational movements of the eye-tracking camera with respect to the subjects' eye.

In Chapter 3, the image formation process of the apparent pupil center and CR is described. The apparent pupil is the image of the actual pupil as seen by the eye-tracking camera through the cornea. It is located 3.05mm way from the vertex of the curvature of the cornea and always falls on the optical axis of the eye, no matter how the eye translates or rotates. The CR is the image of an illumination source formed by the cornea. Because the radius of the cornea is much less than that of the sclera, which is the white portion of the eye, the CR moves at a slower pace than the eye rotation (but in the same direction). Although the CR position changes when the light source moves closer or further away from the cornea, it is always in line with the ray that goes through the corneal center of curvature. For collimated illumination, the CR lies at the intersection point of the ray passing through the corneal center with the corneal focal plane.

Chapter 5 reviews the different kinds of eye tracking techniques and surveys the illumination structures in the state-of-the-art video-based eye tracking systems. Chapter 6 summarizes the key issues in current video-based eye trackers and evaluates several potential illumination configurations for improving the performance of the tracking system. These two chapters, as well as Chapter 3, provide a solid foundation for the work on optimizing video-based eye tracking techniques presented in Chapter 7 and Chapter 8.

In order to separate the translational movements from rotational eye movements, in Chapter 7, the optical relationships between pupil and CR displacements during these two types of movements were derived with the focus on true rotational eye movements within the head about the eye's rotational center and translational eye movements due to relative movements between the eye and the eye-tracking apparatus. The relationships are modelled by two values which have been termed the rotational and translational gains. Because the distance of the illumination source from the eye plays an important role in determining the position of the CR, the gains are derived for the cases of collimated and near-source illuminations. It is shown that the values derived for collimated illumination satisfy the equation derived for near-source illumination for a source at a great distance. Given the fact that the gain values are dependent on the parameters of the eye, they fluctuate with variations in the eyes of individual subjects. The gains from the optical derivations were verified by an experiment that empirically measured the translational and rotational gain values of five individuals. The knowledge of these values presents the possibility of algorithms to improve video-based eye tracking.

In Chapter 8, the rotational gain is used in an algorithm to increase the measurement range and robustness of a new eye tracker design. A structured illumination configuration, utilizing an array of IREDs to illuminate the eye, is proposed to address two common problems in current video-based eye tracking systems: 1) loss of CR during extreme eye movements, and 2) specular reflection artifacts being mistaken for the desired CR. The design provides reliable difference vectors (P-CR) even during extreme eye movements. In the algorithm, the rotational gain is used to predict the location of desired CRs. Potential CRs (both genuine and false) and the pupil are first detected by using global and local statistical information in the eye image. Then genuine CRs are discriminated by selecting

those potential CRs that fall within predicted square regions centered on the CR positions calculated using the eye gain (along with previous CR position and pupil position in current and previous frames). The CR prediction component of this algorithm can be applied to any video-based eye tracking system that tracks the CR. Combining the Structured Illumination with this unique CR prediction technique, the new eye tracking system design is capable of measuring a wide range of eye movements and effectively reduces the threat of spurious reflections being taken as the desired CR, which is a significant problem in video-based eye tracking (especially in daylight).

Another application of the gain values was reported in Kolakowski & Pelz (2006) and Li, Kolakowski, et al. (2008). The rotational and translational gain values were used in an algorithm that separates rotational movements (eye-in-head movements) from translational movements (camera movements with respect to eye) to obtain pupil displacement due only to rotations of the eye within the head.

The cornea is modeled as a sphere in Chapter 7 to derive the offset relationships between the pupil and CR, and the obtained eye gain of 0.5 is used to predict the desired CR position in Section 8.6. This prediction routine could be improved by using an elliptical corneal model (such as those described in Section 3.1). This would generate dynamic ratios according to eye position instead of one fixed value (≈ 0.5) as in the current implementation.

Three coordinate systems are applied in head-mounted video-based eye trackers (e.g., the RIT Eye Tracker with the new Structured Illumination design): head, camera and world (scene or display) coordinates. The rotational and torsional movements of the eye within its socket are commonly defined in head coordinates. In video-based eye tracking, these motions as well as translational eye movements (camera movements) are

captured by video cameras, and the positions of the eye features (the pupil and CR) are used to determine the line of sight. These eye features seen in the image (like those obtained in Chapter 8) are represented in camera coordinates. However, what is of most interest in is where the eye is directed to in the world, so the eye orientations calculated from the positional difference vector between the pupil and CR (see Equation 5-2) need to be converted to world coordinates. Frequently, this coordinate system is also called screen coordinates if a computer display is used to present the stimulus.

The mapping from camera coordinates to world coordinates is a well-known process in video-based eye tracking and can be achieved by a conventional calibration procedure. The subject is instructed to look at a number of points whose positions are known (taking nine calibration points as an example, though five and seventeen points are also common). When the subject fixates on one of nine points, the eye position in the image (in camera coordinates) and the position of the calibration point (in world coordinates) are recorded and then fed into some polynomial function. For a second-order polynomial, the function is defined as:

$$\mathrm{w_x} = c_{x0} + c_{x1}x + c_{x2}y + c_{x3}xy + c_{x4}x^2 + c_{x5}y^2$$

$$\mathrm{w_y} = c_{y0} + c_{y1}x + c_{y2}y + c_{y3}xy + c_{y4}x^2 + c_{y5}y^2$$

where (w_x, w_y) are the position of a calibration point in world coordinates and (x, y) is the difference vector (P-CR) in camera coordinates. There are a total of 12 unknowns ($c_{x0} - c_{x5}\ and\ c_{y0} - c_{y5}$) and 18 equations (two for each calibration point). The problem is over constrained and can be solved by a least-squares approach (Morimoto & Mimica, 2005). The P-CR vector for each calibration point can also be computed as the average of estimates during a fixation period of one point (e.g., 60 estimates/second can be obtained by a 60Hz NTSC camera) (Guestrin & Eizenman, 2006).

The new eye tracking design presented in this dissertation aims to extend the measurement range and reduce the adverse effect of spurious CRs being mistakenly taken as a genuine CR. Although the new design is expected to be comparable with the current system (i.e., RIT Potable Eye Tracker) in many important aspects, a validation study should be conducted to quantitatively evaluate the new eye tracking system. The validation step can be accompanied with artificial eyes, where the ground truth of "eye" motions can be known. The measured accuracy of the system may not represent the true accuracy when measuring real eye movements, however, because the eye is such a complicated system that many influencing factors cannot be taken into account with the artificial eyes (such as the deviation between the optical and visual axes and the eye translation within the socket). Other possible approaches include tracking real observers and a comparison study with other commercial eye tracking systems (e.g., DPI tracker in Section 5.4.1). The metrics for evaluating the system could include measurement range, accuracy within small FOV (e.g., within ±25°) and large FOV (e.g., between ±25° and ±40°), noise level, precision, repeatability, success in tracking different observers, etc.

11.2 Eye-Tracking Technique Applications

The study of the dynamics of eye movements is one of the important applications of video-based eye tracking. Chapter 4 overviews the different types of eye movements and introduces their primary characteristics. Smooth pursuit minimizes or eliminates the retina slips when the eye is tracking a small moving target. In the presented dissertation, two application studies are conducted to explore the performances of pursuit eye movements in two different conditions: (1) a controlled static condition, and (2) an

unconstrained vibrating condition. The two studies are also different in terms of viewing distance to the target, whereas the observer is tracking a target moving at a distance in the first study, the observer in the second study is asked to follow a moving target on a nearby mobile display.

The eye tracking study in the controlled condition in Chapter 9 concludes that the extended targets of realistic images improve the visual performance (i.e., reduced retinal slip) - in comparison with a small target. The upper limit of eye velocity, when tracking a small target in the static laboratory with constrained head and body movements, is found to be 64 deg/sec, while the eye velocity saturates at high velocities when tracking the extended targets. Those findings expand our knowledge about the dynamics of smooth pursuit, one of the most important types of eye movement.

Increasing the angular extent of the stimulus from the dot to the apple image, which was large enough to cover the fovea, caused a significant increase in smooth pursuit gain. Increasing the image size beyond the fovea (from the apple to the flower image), however, did not lead to further significant increases in gain. It is hypothesized that the motion signals activated in the parafovea and peripheral retina are summed with (or weighted-averaged with those triggered in the central fovea. Further work with stimuli of different angular extent is needed to investigate this hypothesis.

In Chapter 10, the performance of smooth pursuit eye movements was found to be greatly diminished when the observer was tracking a target in a moving vehicle, even with unconstrained head movements. The free head motion in the experiment did tend to enhance the visual performance, but the pursuit tracking is not smooth, having significant retinal slip, even when the observer is following a slowly moving object in the vibrating condition. With the increasing popularity of mobile devices, this work lays the

foundation for understanding the different gaze behaviors when the observer is performing visual tasks on a small display in a commuting vehicle when compared to those in a static laboratory.

Further studies, such as free head movements in the static condition or constrained head movements in the vibrating condition, may help determine the effect of head movement on performance. Studies of tracking the targets on the mobile display in a static condition, for example, and the subject doing the same tracking task when sitting in a stationary car would be beneficial to learn more about how the vibration in the vehicle affects the visual performance of target pursuit.

Bibliography

Abel, L. A., & Hertle, R. W. (1988). Effects of Psychoactive Drugs on Ocular Motor Behavior. In C. W. Johnston & F. J. Pirozzolo (Eds.), *Neuropsychology of Eye Movements*. Lawrence Erlbaum Associates: Hillsdale, NJ.

Alex, C., Myron, F., Jianchang, M., et al. (2001). *A Comparison of Classifiers for Real-Time Eye Detection*. Paper presented at the Proceedings of the International Conference on Artificial Neural Networks.

ASL. (2005). *Eye Tracking System Instructions, ASL Eye-Trac 6000, Head Mounted Optics.*: Applied Science Laboratories.

Aslin, R. N., & McMurray, B. (2004). Automated Corneal-Reflection Eye Tracking in Infancy: Methodological Developments and Applications to Cognition. *Infancy, 6*(2), 155-163.

Atchison, D. A., & Smith, G. (2000). *Optics of the Human Eye* (1st ed.): Butterworth-Heinemann.

Babcock, J., & Pelz, J. (2004). *Building a lightweight eyetracking headgear.* Paper presented at the Proceedings of the symposium on eye tracking research & applications, San Antonio, Texas.

Bahill, A. T., Clark, M., & Stark, L. (1975). The Main Sequence, a tool for studying human eye movements. *Mathematical Biosciences, 24*, 191-204.

Barker, T. B. (2005). *Quality by Experimental Design* (3rd ed.): Chapman & Hall/CRC.

Beutter, B. R., Borthwick, R., & Stone, L. S. (1998). *Optimizing infra-red light collection for high-speed video-based eye tracking*: The Vision Science and Technology Group at NASA Ames Research Center.

Borah, J. (1989). *Helmet Mounted Eye tracking for Virtual Panoramic Display Systems-Volumes I*: Wright-Patterson Air Force Base, OH: Harry G. Armstrong Aerospace Medical Research Laboratory, Air Force Systems Command.

Cann, A., Salmoni, A., & Eger, T. (2004). Predictors of whole-body vibration exposure experienced by highway transport truck operators *Ergonomics, 47*(13), 1432-1453.

Carpenter, R. H. S. (1988). *Movements of the Eyes*. London: Pion Limited.

Chen, J., Chang, W., Shih, T., et al. (2003). Predictors of whole-body vibration levels among urban taxi drivers. *Ergonomics, 46*, 1075-1090.

Chronos-Vision. CHRONOS VISION GmbH, Wiesenweg 9, 12247 Berlin, Germany. www.chronos-vision.de.

Clarkson, T. G. (1989). Safety aspects in the use of infrared detection systems. *International Journal of Electronics [INT. J. ELECTRON.]. , 66*(6), 929-934.

Cline, D., Hofstetter, H., & Griffin, J. (1989). *Dictionary of Visual Science* (4th ed.). Radnor, Pennsylvania: Chilton Trade Book Publishing.

Collewijn, H. (1991). The Optokinetic Contribution. In R. H. S. Carpenter (Ed.), *Vision and Visual Dysfunction; Volume 8: Eye Movements* (pp. 45-70): MacMillan, London

Collewijn, H. (1999). Eye Movement Recording. In R. H. S. Carpenter & J. G. Robson (Eds.), *Vision research; A Practical Guide to Laboratory Methods* (pp. 245-285): Oxford University Press.

Collewijn, H., Steinman, R., & van der Steen, J. (1985). The performance of the smooth pursuit eye movement system during passive and self-generated stimulus motion. *Journal of Physiology (London), 366*, 19p.

Collewijn, H., & Tamminga, E. P. (1984). Human smooth and saccadic eye movements during voluntary pursuit of different target motions on different backgrounds.

Journal of Physiology, 351, 217-250.

Cornsweet, T. N., & Crane, H. D. (1973). Accurate two-dimensional eye tracker using first and fourth Purkinje images. *J Opt Soc Am, 63*(8), 921-928.

Crane., H. D., & Steele, C. M. (1985). Generation-V dual-Purkinje-image eyetracker. *Applied Optics, 24*(4), 527-537.

Daugman, J. (2004). How iris recognition works. *Circuits and Systems for Video Technology, IEEE Transactions on, 14*(1), 21-30.

Davson, H. (1962). *The Eye, Volume 4: Visual Optics and the Optical Space Sense* (Vol. Volume 2). New York: Academic Press.

Davson, H. (1990). *Physiology of the Eye* (5th ed.): Palgrave Macmillan.

Delabarre, E. B. (1898). A Method of Recording Eye-Movements. *American Journal of Psychology, 9*, 572-574.

Deubel, H., & Bridgeman, B. (1995a). Fourth Purkinje image signals reveal eye-lens deviations and retinal image distortions during saccades. *Vision Res, 35*(4), 529-538.

Deubel, H., & Bridgeman, B. (1995b). Perceptual consequences of ocular lens overshoot during saccadic eye movements. *Vision Res, 35*(20), 2897-2902.

DiScenna, A. O., Das, V., Zivotofsky, A. Z., et al. (1995). Evaluation of a video tracking device for measurement of horizontal and vertical eye rotations during locomotion. *J Neurosci Methods, 58*(1-2), 89-94.

Dickman, J. D. (2006). The Vestibular System. In D. E. Haines (Ed.), *Fundamental Neuroscience* (3rd Edition ed., pp. 341–358): Williams & Wilkins Co.

Duchowski, A. T. (2002). *Eye Tracking Methodology: Theory and Practice.*: Springer.

Ebisawa, Y. (1995). Unconstrained pupil detection technique using two light sources and the image difference method. *Visualization and Intelligent Design in engineering and architecture*, 79-89.

El-Mar. EL-MAR INC., from 48 Evanston Drive, Downsview, Ontario, Canada, M3H 5P3.: http://home.interlog.com/~elmarinc/menu.htm

EMED. Eye Movement Equipment Database (EMED). from Applied Vision Research Centre (AVRC), Loughborough University: http://www.lboro.ac.uk/research/esri/applied-vision/projects/vision_resources/emed.htm

ERT. Eye Response Technologies. from 100 2nd Street, NW Charlottesville, VA 22902, USA.: www.eyeresponse.com

Ferwerda, J. A. (1998). *Fundamentals of spatial vision*.

Fitzgibbon, A., Pilu, M., & Fisher, R. B. (1999). Direct least square fitting of ellipses. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 21*(5), 476-480.

Fourward. Fourward Technologies, Inc., from 2464 Maple Ave. Buena Vista, VA 24416,USA, www.fourward.com. Accessed on May 8th, 2007.:

Frens, M. A., & van der Geest, J. N. (2002). Scleral search coils influence saccade dynamics. *J Neurophysiol, 88*(2), 692-698.

Gitelman, D. R. (2002). ILAB: a program for postexperimental eye movement analysis. *Behav Res Methods Instrum Comput, 34*(4), 605-612.

Glenny, G., & Heywood, S. (1979). Hans Gertz revisited: the different effects of invisibility and darkness on pursuit eye movements. *Perception, 8*(1), 31-36.

Glenstrup, A. J., & Engell-Nielsen, T. (1995). *Eye Controlled Media: Present and Future State.* University of Copenhagen DIKU Universitetsparken 1 DK-2100 Denmark.

Grand, Y. L., & Hage, S. G. E. (1980). *Physiological Optics*. Berlin, Heidelberg, New York: Springer-Verlag

Green, P. (1992). *Review of eye fixation recording methods and equipment* (No. Technical Report UTMTRI-92-28). Ann Arbor, Michigan: The University of Michigan,Transportation Research Institute.

Greg, W., Gary, B., Leandra, V., et al. (1999). *The HiBall Tracker: high-performance wide-area tracking for virtual and augmented environments*. Paper presented at the Proceedings of the ACM symposium on Virtual reality software and technology.

Griffin, M. J. (1975). Levels of Whole-body Vibration Affecting Human Vision. *Aviation, Space and Environmental Medicine, 46*(8), 1033-1040.

Griffin, M. J. (1981). Biodynamic Response To Whole-Body Vibration. *The Shock and Vibration Digest, 13*(8), 3-12.

Griffin, M. J. (1992). Vibration. In D. M. Jones & A. P. Smith (Eds.), *Handbook of Human Performance, Vol. 1: The Physical Environment*: London: Academic Press.

Griffin, M. J., & Lewis, C. H. (1978). A review of the effects of vibration on visual acuity and continuous manual control, part I: visual acuity. *Journal of Sound and Vibration, 56*(3), 383–413.

Guestrin, E. D., & Eizenman, M. (2006). General theory of remote gaze estimation using the pupil center and corneal reflections. *Biomedical Engineering, IEEE Transactions on, 53*(6), 1124-1133.

Hallett, P. E. (1986). Eye movements. In K. R. Boff, L. Kaufman & J. P. Thomas (Eds.), *Handbook of Perception and Human Performance* (Vol. 1, pp. chapter 10). New York: Wiley.

Haro, A., Flickner, M., & Essa, I. (2000). *Detecting and tracking eyes by using their physiological properties, dynamics, and appearance.* Paper presented at the Computer Vision and Pattern Recognition.

Hashiba, M., Hattori, T., Watabe, H., et al. (1996). Influence of superimposed optokinetic stimulus on smooth pursuit eye movements elicited by sinusoidal spot target oscillation. *Acta oto-laryngologica. Supplementum, 525*, 167-171.

Hayhoe, M. M. (2004a). Advances in relating eye movements and cognition. *Infancy, 6*(2), 267-274.

Hayhoe, M. M., and Ballard, D. H. (2004b). Eye Movements in Natural Behavior. *Trends in Cognitive Science, 9*(4).

Hecht, E. (1998). *Optics* Addison Wesley; 3rd edition, page155-159.

Heinen, S. J., & Keller, E. L. (2004). Smooth pursuit eye movements: Recent advances. In L. Chulupa & J. Warner (Eds.), *The Visual Neurosciences* (pp. 1402-1414): MIT Press: Cambridge.

Heinen, S. J., & Watamaniuk, S. N. J. (1998). Spatial integration in human smooth pursuit. *Vision Research, 38*(23), 3785-3794.

Hua, H., Krishnaswamy, P., & Rolland, J. P. (2006). Video-based eyetracking methods and algorithms in head-mounted displays. *Optics Express, 14*(10), 4328-4350

Ilg, U. J. (1997). Slow eye movements. *Progress in neurobiology, 53*(3), 293-329.

Irving, E. L., Zacher, J. E., Allison, R. S., et al. (2003). Effects of scleral search coil wear on visual function. *Invest Ophthalmol Vis Sci, 44*(5), 1933-1938.

ISCAN. ISCAN, Inc., from 89 Cambridge Street, Burlington, MA 01803, U.S.A., www.iscaninc.com/:

ISCAN. (1996). *Operating Instructions*: ISCAN, Inc. 89 Cambridge Street, Burlington, MA 01803, U.S.A. .

Johanning, E., Wilder, D. G., Landrigan, P. J., et al. (1991). Whole-Body Vibration

Exposure in Subway Cars and Review of Adverse Health Effects. *Journal of Occupational Medicine, 33*(5).

Judge, J., Caravolas, M., & Knox, P. C. (2006). Smooth pursuit eye movements and phonological processing in adults with dyslexia. *Cognitive Neuropsychology, 23*(8), 1174 -1189.

Juhola, M., Jantti, V., Pyykko, I., et al. (1985). Detection of saccadic eye movements using a non-recursive adaptive digital filter. *Comput Methods Programs Biomed, 21*(2), 81-88.

Karlene, N., Cindy, W., David, K., et al. (2002). *Differences in the infrared bright pupil response of human eyes*. Paper presented at the Proceedings of the symposium on Eye tracking research \& applications.

Karn, K. S. (2000). *"Saccade pickers" vs. "fixation pickers": the effect of eye tracking instrumentation on research*. Paper presented at the Proceedings of the symposium on Eye tracking research & applications.

Kolakowski, S. M., & Pelz, J. B. (2006). *Compensating for eye tracker camera movement*. Paper presented at the Proceedings of the 2006 symposium on eye tracking research & applications.

Kolb, H. (2003). Structure of the retina: American Scientist.

Kowler, E. (1989). Cognitive expectations, not habits, control anticipatory smooth oculomotor pursuit. *Vision Res, 29*(9), 1049-1057.

Kundur, D., & Hatzinakos, D. (1996). Blind Image Deconvolution Revisited. *IEEE Signal Processing Magazine, 13*(3), 43-64.

Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception, 28*(11), 1311-1328.

Land, M. F., & Lee, D. N. (1994). Where we look when we steer. *Nature, 369*(6483), 742-744.

Land, M. F., & McLeod, P. (2000). From eye movements to actions: how batsmen hit the ball. *Nat Neurosci, 3*(12), 1340-1345.

LCT. LC Technologies. from 3955 Pender Drive, Suite 120, Fairfax, Virginia 22030 U.S.A., www.eyegaze.com/:

Lewis, C. H., & Griffin, M. J. (1980). Predicting the effects of vibration frequency and axis and seating condition on the reading of numeric displays. *Ergonomics, 23*(5), 485-501.

Li, D. (2006). *Low-cost eye-tracking for human computer interation.* Iowa State University, Ames, Iowa.

Li, D., & Parkhurst, D. J. (2006). openEyes: an open-hardware open-source system for low-cost eye tracking. *Journal of Modern Optics, 53*, 1295-1311.

Li, F., Kolakowski, S., & Pelz, J. (2008). A model-based approach to video eye tracking. *Journal of Modern Optics, 55*(4 & 5), 503 - 531.

Lipps, M. (2004). *Task influence of scene content selected by active vision.* Rochester Institute of Technology, Rochester.

Masson, G., Proteau, L., & Mestre, D. R. (1995). Effects of stationary and moving textured backgrounds on the visuo-oculo-manual tracking in humans. *Vision Research, 35*(6), 837-852.

Merchant, J., Morrissette, R., & Porterfield, J. L. (1974). Remote measurement of eye direction allowing subject motion over one cubic foot of space. *IEEE Trans Biomed Eng, 21*(4), 309-317.

Meyer, C. H., Lasker, A. G., & Robinson, D. A. (1985). The upper limit of human

smooth pursuit velocity. *Vision Research, 25*(4), 561-563.

Millodot, M. (1972). Variation of visual acuity in the central region of the retina. *The British journal of physiological optics, 27*(1), 24-28.

Moore, S. (1999). *Technical Issues Facing VOG Development.* Paper presented at the Workshop on the coordination of developments of video systems for the measurement of 3-dimensional eye movements, Tübingen, Germany (Nov 30 - Dec 2, 1999).

Morimoto, C. H., Koons, D., Amir, A., et al. (1999). *Frame-rate pupil detector and gaze tracker.* Paper presented at the IEEE ICCV'99 Frame-Rate Workshop.

Morimoto, C. H., Koons, D., Amir, A., et al. (2000). Pupil detection and tracking using multiple light sources. *Image and Vision Computing, 18*(4), 331-335

Morimoto, C. H., & Mimica, M. R. M. (2005). Eye gaze tracking techniques for interactive applications (Vol. 98, pp. 4-24): Elsevier Science Inc.

Moseley, M. J., & Griffin, M. J. (1986). Effects of display vibration and whole-body vibration on visual performance. *Ergonomics, 29*, 977–983.

Munn, S. M. (2009). An algorithm for identifying and tagging fixations to simplify the analysis of data collected by portable eye trackers. *ACM Transactions on Applied Perception (TAP), 6*(3).

NASA. (2002). *High-Frame-Rate CCD Camera Having Subwindow Capability* (No. NPO-30564).

NASA. (2004). *Event-Driven Random-Access-Windowing CCD Imaging System* (No. NPO-30878).

Niemann, T., & Hoffmann, K. P. (1997). The influence of stationary and moving textured backgrounds on smooth-pursuit initiation and steady state pursuit in humans.

Experimental brain research, 115(3), 531-540.

Pansky, B. (1988). *Review of Neuroscience*: MacMillan Publishing Company; 2nd Ed

Pedrotti, L. S., & Pedrotti, F. L. (1998). *Optics and Vision*: Prentice-Hall, Inc.

Pelz, J. B. (2004). Portable eyetracking in natural behavior. *Journal of Vision, 4*(11), 14-14.

Pelz, J. B., & Canosa, R. (2001). Oculomotor behavior and perceptual strategies in complex tasks. *Vision Res, 41*(25-26), 3587-3596.

Pola, J., & Wyatt, H. J. (1985). Active and passive smooth eye movements: effects of stimulus size and location. *Vision Research, 25*(8), 1063-1076.

Pola, J., & Wyatt, H. J. (1991). The smooth pursuit eye movement system: response characteristics, stimuli, and mechanisms. In R. H. S. Carpenter (Ed.), *Vision and Visual Dysfunction; Volume 8: Eye Movements* (pp. 138-156): MacMillan, London

R.Tole., J., & R.Young, L. (1980). Digital filters for saccade and fixation detection. In D. F. Fisher, R. A. Monty & J. W. Senders (Eds.), *Eye Movements: Cognition and Visual Perception* (pp. 7-17). Hillsdale, NJ: Lawrence Erlbaum Associates.

Rashbass, C. (1961). The relationship between saccadic and smooth tracking eye movements. *Journal of Physiology, 159*, 326-338.

Reeves, P. (1920). The response of the average pupil to various intensities of light *Journal of the Optical Society of America, 4*(2), 35.

Richardson, D. C., & Spivey, M. J. (2004). Eye Tracking: Characteristics and Methods In G. Wnek & G. Bowlin (Eds.), *Encyclopedia of Biomaterials and Biomedical Engineering* (pp. 568-572). New York: Marcel Dekker, Inc.

Robinson, D. A. (1965). The mechanics of human smooth pursuit eye movement.

Journal of Physiology, 180(3), 569-591.

Ross, R. G., Olincy, A., Harris, J. G., et al. (1999). The effects of age on a smooth pursuit tracking task in adults with schizophrenia and normal subjects. *Biological Psychiatry, 46*(3), 383-391.

Salvucci, D. D., & Goldberg, J. H. (2000). *Identifying fixations and saccades in eye-tracking protocols*. Paper presented at the Proceedings of the symposium on Eye tracking research \& applications.

Schlenker, R., Rudolf, C., Berg, P., et al. (1994). Smooth-pursuit eye movement dysfunction in schizophrenia: the role of attention and general psychomotor dysfunctions. *European Archives of Psychiatry and Clinical Neuroscience, 244*(3), 153-160.

Siedlecki, D., Kasprzak, H., & Pierscionek, B. K. (1997). Schematic eye with a gradient-index lens and aspheric surfaces *Optics Letters, 29*(11), 1197-1199.

Sliney, D., & Wolbarst, M. (1981). *Safety with Lasers and Other Optical Sources*: Plenum Press.

SMI. SensoMotoric Instruments GmbH. from Warthestraße 21, D-14513 Teltow/Berlin, Germany, www.smi.de:

Spering, M., Kerzel, D., Braun, D. I., et al. (2005). Effects of contrast on smooth pursuit eye movements. *Journal of Vision, 5*(5), 455-465.

SR_Research. SR Research Ltd., from 5516 Main St., Osgoode ON, Canada K0A 2W0. : www.eyelinkinfo.com/

SR_Research. (2002). *EyeLink II User Manual, Version 1.05*: SR Research Ltd, Mississauga, Canada.

Stott, J. R. R. (1984). The vertical vestibulo-ocular reflex and ocular resonance. *Vision*

Research, 24, 949-960.

Talukder, A., Morookian, J. M., Monacos, S., et al. (2005). Eye-tracking architecture for biometrics and remote monitoring. *Applied Optics, 44*(5), 693-700.

Tan, K.-H., Kriegman, D. J., & Ahuja, N. (2002). *Appearance-based eye gaze estimation*.

Triesch, J., Sullivan, B. T., Hayhoe, M. M., et al. (2002). *Saccade contingent updating in virtual reality*. Paper presented at the Proceedings of the symposium on Eye tracking research \& applications.

van den Berg, A. V., & Collewijn, H. (1986). Human smooth pursuit: Effects of stimulus extent and of spatial and temporal constraints of the pursuit trajectory. *Vision Research, 26*(8), 1209-1222.

W.N.Charman. (1994). Chapter 24: Optics of the Eye *Handbook of Optics* McGraw-Hill Professional; 2 edition

Walsh, G. (1988). The effect of mydriasis on the pupillary centration of the human eye. *Ophthalmic Physiol Opt, 8*(2), 178-182.

Watanabe, J., Ando, H., Sekiguchi, D., et al. (2003). *The study of remote saccade sensing system based on retroreflective feature of the retina.* Paper presented at the Proceedings of the 13th International Conference on Artificial Reality and Telexistence, Keio University, Tokyo, Japan.

Welch, G., Bishop, G., Vicci, L., et al. (2001). *High-Performance Wide-Area Optical Tracking -The HiBall Tracking System.* Paper presented at the Teleoperators and Virtual Environments

Westheimer, G. (1954). Eye movement responses to a horizontally moving visual stimulus. *A.M.A. archives of ophthalmology, 52*(6), 932-941.

White, O. B., Saint-Cyr, J. A., Tomlinson, R. D., et al. (1983). Ocular motor deficits in

Parkinson's disease. II. Control of the saccadic and smooth pursuit systems. *Brain, 106*(pt 3), 571-587.

Wilson, M. A., Campbell, M. C., & Simonet, P. (1992). Change of pupil centration with change of illumination and pupil size. *Optom Vis Sci, 69*(2), 129-136.

Yang, Y., Thompson, K., & Burns, S. A. (2002). Pupil location under mesopic, photopic, and pharmacologically dilated conditions. *Invest Ophthalmol Vis Sci, 43*(7), 2508-2512.

Yarbus, A. L. (1967). *Eye Movements and Vision* New York: Plenum.

Yee, R. D., Daniels, S. A., Jones, O. W., et al. (1983). Effects of an optokinetic background on pursuit eye movements. *Investigative Ophthalmology & Visual Science, 24*(8), 1115-1122.

Young, L. R., & Sheena, D. (1975). Survey of eye movement recording methods. *Behaviour Research Methods & Instrumentation, 7*(5), 397-429.

Zambarbieri, D., Schmid, R., Prablanc, C., et al. (1981). Characteristics of eye movements evoked by the presentation of acoustic targets. In A. F. Fuchs & W. Becker (Eds.), *Progress in Oculomotor Research* (pp. 559-566): Elsevier/North Holland, New York.

Zhu, Z., & Ji, Q. (2005). Robust real-time eye detection and tracking under variable lighting conditions and various face orientations. *Computer Vision and Image Understanding, 98*(1), 124-154.